

КАЧЕСТВО ПОИСКА И ПОИСКОВЫЕ СИСТЕМЫ В РУНЕТЕ

АНАЛИТИЧЕСКИЙ ОБЗОР,
2014 — I КВАРТАЛ 2015 ГОДА

АВТОРЫ ОТЧЕТА:

ЛЮДМИЛА КУДРЯВЦЕВА, СЕРГЕЙ ВАЛЬКОВСКИЙ,
МИХАИЛ ВОЛОВИЧ, КОНСТАНТИН РОЩУПКИН



Ашманов
и партнеры

СОДЕРЖАНИЕ

1. О компании «Ашманов и партнеры»	4
--	---

ЧАСТЬ 1

2. Качество поиска: критерии и динамика	5
Независимая оценка качества поиска.....	5
Что нового на Analyzethis.ru появилось в 2014 году	6
Как работают анализаторы качества поиска.....	7
3. Качество поиска по навигационным запросам	8
Анализатор качества навигационного поиска.....	9
Анализатор качества поиска периферийных сайтов	10
Анализатор качества внутрисайтовой навигации.....	11
Анализатор качества поиска персон.....	12
Анализатор качества поиска персональных блогов	13
Анализатор качества навигационного поиска в социальных сетях.....	14
4. Качество поиска по информационным запросам	15
Анализатор качества поиска цитат.....	16
Анализатор качества поиска крылатых слов.....	17
Анализатор качества поиска ответов.....	18
Анализатор качества поиска оригиналов новостей.....	19
Анализатор качества поиска адресов	20
5. Качество поиска по транзакционным запросам	21
Анализатор качества поиска официальных версий программного обеспечения	22
Анализатор качества поиска официальных версий контента.....	23
6. Актуальность результатов поиска	24
Анализатор актуальности поиска: должности	25
Анализатор актуальности поиска: телефоны	26
Анализатор скорости индексации.....	27
7. Полнота и разнообразие результатов поиска	28
Анализатор полноты индекса.....	29
Анализатор качества тематического поиска.....	30
Анализатор качества поиска омонимов	31
8. Поиск из регионов	32
Анализатор регионального поиска	33
Анализатор качества регионального навигационного поиска	34

9.	Понимание запроса.....	35
	Анализатор качества подсказок.....	36
	Анализатор устойчивости к опечаткам в запросе.....	37
	Анализатор запросов-синонимов.....	38
	Анализатор поиска по запросам с числами.....	39
	Анализатор поиска моделей.....	40
10.	Анализаторы ошибок.....	41
	Анализатор замены редких слов.....	42
	Анализатор замены редких фраз.....	43
	Анализатор неразрывных словосочетаний.....	44
	Анализатор поиска имен.....	45
	Анализатор грамматики.....	46
11.	Раздражающие факторы.....	47
	Анализатор спама в поисковой выдаче.....	48
	Анализатор «сайтов для взрослых» в поисковой выдаче.....	49
	Анализатор качества семейного фильтра.....	50
	Анализатор рекламной нагрузки.....	51
	Анализатор порнорекламы.....	52
	Анализатор вирусных угроз.....	53
12.	Общие характеристики поиска.....	54
	Ассессорский анализатор.....	55
	Качество снипетов.....	56
	Анализатор апдейтов.....	57
	Анализатор скоростирегионального поиска.....	58

ЧАСТЬ 2

1. Основные события на рынке поиска в Рунете.....	59
1.1. Эволюция численности, состава и поведения российской интернет-аудитории	59
1.2. Рост популярности мобильных устройств	60
1.3. Усиление государственного регулирования.....	60
1.4. Неблагоприятная экономическая ситуация.....	60
1.5. Конкуренция и сотрудничество с социальными сетями	61
2. Яндекс в 2014 — I квартале 2015 года	62
2.1. Борьба с Google за мобильную аудиторию.....	62
2.2. Ужесточение борьбы с манипуляциями поиском.....	63
2.3. Развитие альтернативных моделей получения прибыли, работа с большими данными.....	64
2.4. Заморозка или перезапуск «Островов»	64
2.5. Акцент на рекомендательные системы и технологии персонализации.....	65
2.6. Закрытие сервисов и покупка компаний.....	65
2.7. Яндекс в массовой культуре	65
3. Google в 2014 — I квартале 2015 года	66
3.1. Ранжирование сайтов на основе достоверности.....	66
3.2. Курс на адаптацию всех сайтов под мобильные устройства.....	66
3.3. Отказ от развития социальной сети «классического» типа и интеграции ее с поиском.....	66
4. Поиск@Mail.ru в 2014 — I квартале 2015 года.....	67
5. «Спутник» в 2014 — I квартале 2015 года	68

ЧАСТЬ 3

1. Новый проект: SEO-анализатор	69
Что такое коммерческие факторы ранжирования.....	69
По результатам наших наблюдений можно достаточно уверенно сделать два вывода.....	70
Разумеется, исследования Лаборатории не ограничиваются коммерческими факторами.....	74

О КОМПАНИИ «АШМАНОВ И ПАРТНЕРЫ»

4

Ежегодный аналитический отчет по качеству работы поисковых систем подготовлен компанией «Ашманов и партнеры». Материал основан на данных, полученных при помощи семейства анализаторов проекта AnalyzeThis.

Анализаторы, разработанные компанией, позволяют с высокой точностью оценивать, как с течением времени меняется качество работы основных поисковых систем русскоязычного сегмента Интернета.

Создание этого уникального проекта стало возможным благодаря привлечению к работе лучших специалистов в области поисковых технологий и лингвистики.

Среди прочих разработок компании — сервисы, обеспечивающие высококачественный семантический анализ контента. Они базируются на технологии «Семантическое зеркало», которая также используется системой мониторинга социальных сетей «Крибрум».

Компания «Ашманов и партнеры» является одним из лидеров российского рынка интернет-маркетинга. С 2001 года компания оказывает услуги по комплексному продвижению брендов в Интернете, консалтингу и аналитике.

ЧАСТЬ 1

1.

КАЧЕСТВО ПОИСКА: КРИТЕРИИ И ДИНАМИКА

5

НЕЗАВИСИМАЯ ОЦЕНКА КАЧЕСТВА ПОИСКА

В 2007 году лингвисты и разработчики компании «Ашманов и партнеры» создали методику, позволяющую автоматически оценивать поисковые системы по различным характеристикам. В том же году были запущены первые анализаторы качества поиска на основе технологии сайтов-маркеров.

Срезультатом работы анализаторов можно ознакомиться на сайте проекта AnalyzeThis (www.analyzethis.ru). На сегодня эти анализаторы — единственный независимый тест качества поиска в Рунете. В 2010 году была запущена английская версия анализаторов.

В 2012 году сайт проекта был обновлен, анализаторы были разбиты на группы, что сделало просмотр результатов более удобным.

Сегодня на AnalyzeThis ежедневно публикуются:

- ▶ результаты 40 анализаторов различных «частных» характеристик поиска, оцениваемых автоматически;
- ▶ интегральный показатель качества поиска на основе этих анализаторов;
- ▶ анализаторы, в которых качество поиска оценивается ассессорами по сбалансированному набору запросов и снипетов;
- ▶ количество переходов с результатов поиска на сайты по данным Liveinternet (в том числе с поправкой на процент сайтов со счетчиками Liveinternet).

6 В данном отчете изложены основные события и новости в жизни поисковых машин Рунета и описаны изменения в различных аспектах качества поиска за 2014 год и первый квартал 2015 года.

ЧТО НОВОГО НА ANALYZETHIS.RU ПОЯВИЛОСЬ В 2014 ГОДУ

В 2014 году на Analyzethis.ru появился новый анализатор «коммерческих апдейтов», а анализатор SEO-прессинга был заморожен.

В начале 2014 года в дополнение к анализатору апдейтов, отслеживающему ежедневные изменения в поисковой выдаче, был создан анализатор «коммерческих апдейтов», отслеживающий такие изменения в выдаче по коммерческим запросам. Нововведение было связано с анонсированным в декабре 2013 года Яндексом отказом от учета ссылок при ранжировании поисковой выдачи по коммерческим запросам. Нам стало интересно проследить, как эти изменения повлияют на результаты поиска, и мы создали анализатор, запросы которого очевидно коммерческие, а при продвижении сайтов по этим запросам вероятно активное использование ссылок. Новый анализатор исследует выдачу в различных регионах России, что важно, потому что Яндекс обещал для начала отказаться от использования ссылок в ранжировании только московской выдачи.

Одновременно с открытием анализатора коммерческих апдейтов произошло закрытие анализатора SEO-прессинга. Этот анализатор показывал долю активно продвигаемых коммерческих сайтов узкой направленности в выдаче по достаточно широким запросам. Мы отметили, что все поисковые машины научились успешно справляться с таким прессингом и за шесть лет довели долю таких сайтов с необо-снованных 60-70% до приемлемых 20-25%.

История всех данных анализатора SEO-прессинга сохраняется на сайте, но с марта 2014 года сбор данных для этого анализатора прекращен.

КАК РАБОТАЮТ АНАЛИЗАТОРЫ КАЧЕСТВА ПОИСКА

Каждый анализатор оценивает один из аспектов качества поиска. Большинство анализаторов работает в автоматическом режиме, хотя отдельные анализаторы требуют дополнительной ручной обработки. Анализатор задает всем поисковикам специально подобранный специалистами компании набор запросов, к которым почти всегда привязан набор маркеров, по присутствию (или отсутствию) которых в результатах поиска и оценивает выдачу робот. Маркером может быть конкретный сайт, страница или даже словосочетание — это зависит от того, что именно оценивает анализатор. Поисковая выдача по нескольким тысячам запросов фиксируется и оценивается ежедневно.

Подробное описание работы анализаторов вы можете найти на сайте www.analyzethis.ru или в виде приложения к первому аналитическому отчету «Поиск и контекстная реклама в Рунете 2008», который можно скачать на сайте www.ashmanov.com.

Все результаты работы анализаторов с момента их публикации всегда можно посмотреть на сайте www.analyzethis.ru; во многих случаях статистика доступна за 5 и более лет.

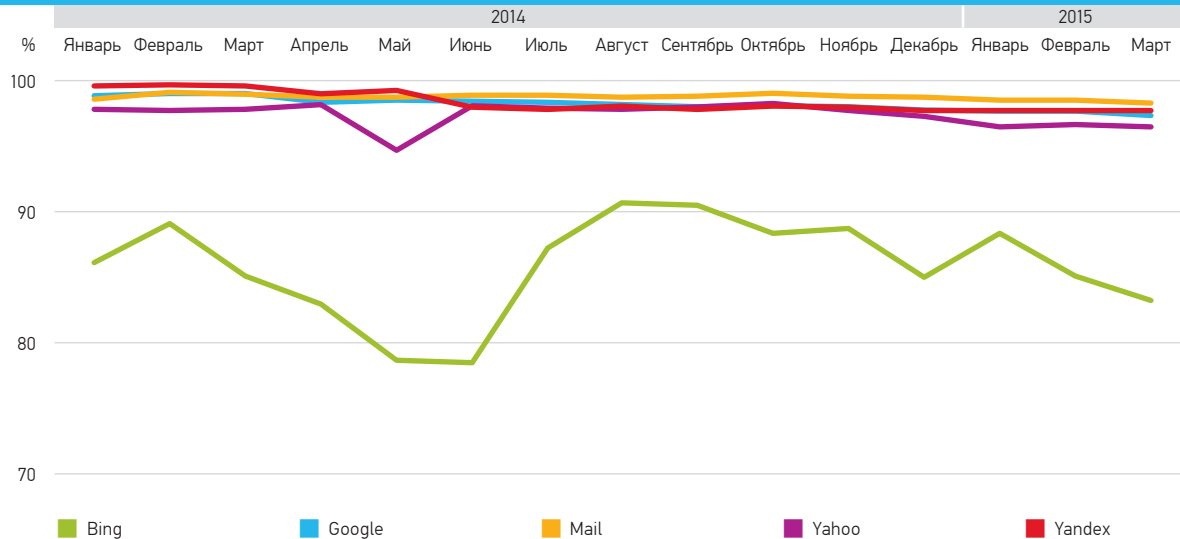
2.

КАЧЕСТВО ПОИСКА ПО НАВИГАЦИОН- НЫМ ЗАПРОСАМ

8

ПОИСК ПО НАВИГАЦИОННЫМ ЗАПРОСАМ — ЭТО ПОИСК ОПРЕДЕЛЕННЫХ СТРАНИЦ И САЙТОВ В ИНТЕРНЕТЕ. НАПРИМЕР, ПО ЗАПРОСУ «АВИТО РУ» ПОЛЬЗОВАТЕЛЬ, СКОРЕЕ ВСЕГО, ХОЧЕТ ПОЛУЧИТЬ ССЫЛКУ НА САЙТ БЕСПЛАТНЫХ ОБЪЯВЛЕНИЙ WWW.AVITO.RU. С ПОИСКОМ ПО ОБЫЧНЫМ НАВИГАЦИОННЫМ ЗАПРОСАМ ВСЕ ПОИСКОВИКИ УЖЕ ДАВНО ОТЛИЧНО СПРАВЛЯЮТСЯ. ИНТЕРЕСНО ПОСМОТРЕТЬ ЧУТЬ ГЛУБЖЕ — НАПРИМЕР, НА ПОИСК ПО РЕГИОНАЛЬНЫМ НАВИГАЦИОННЫМ ЗАПРОСАМ ИЛИ НА ВНУТРИСАЙТОВУЮ НАВИГАЦИЮ.

Рисунок 1. Качество навигационного поиска

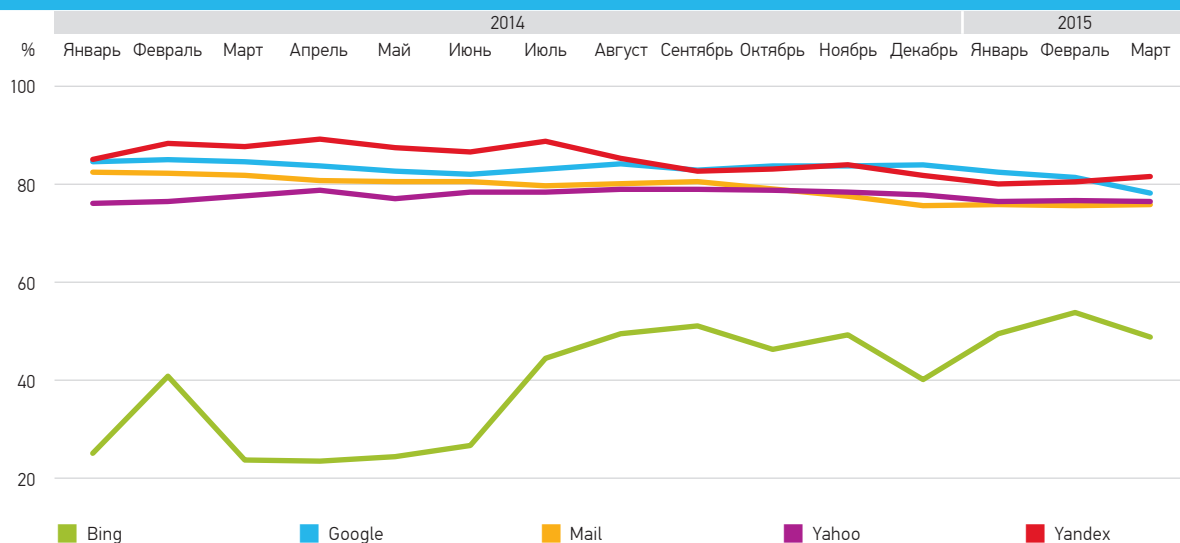


АНАЛИЗАТОР КАЧЕСТВА НАВИГАЦИОННОГО ПОИСКА

Качество навигационного поиска уже давно во всех поисковиках приблизилось к 99%. В конце 2013 года был полностью заменен набор запросов этого анализатора: количество запросов было увеличено в три раза, а сами запросы стали сложнее для поисковиков. Так, новый набор содержит запросы с опечатками, неофициальные названия, названия сайтов в русской транскрипции и т. п. Несмотря на это, в 2014 и начале 2015 года результаты поисковиков оставались стабильно высокими, хотя и немного отошли от максимальных уровней (рис. 1).

Здесь необходимо отметить, что, делая все выводы на основе данных наших анализаторов, мы вынуждены исключить из рассмотрения поисковую машину Bing, т.к. результаты, получаемые от нее анализаторами, слишком сильно отличаются от результатов, которые видит пользователь — Bing выдает по анализаторским запросам ухудшенную «быструю» выдачу. Неплохим источником информации о поиске Bing является поиск Yahoo, использующий поисковый движок Bing, но говорить о полном совпадении было бы неправильно. К динамике Bing на представленных в отчете графиках стоит тоже отнестись критически.

Рисунок 2. Качество поиска периферийных сайтов



10

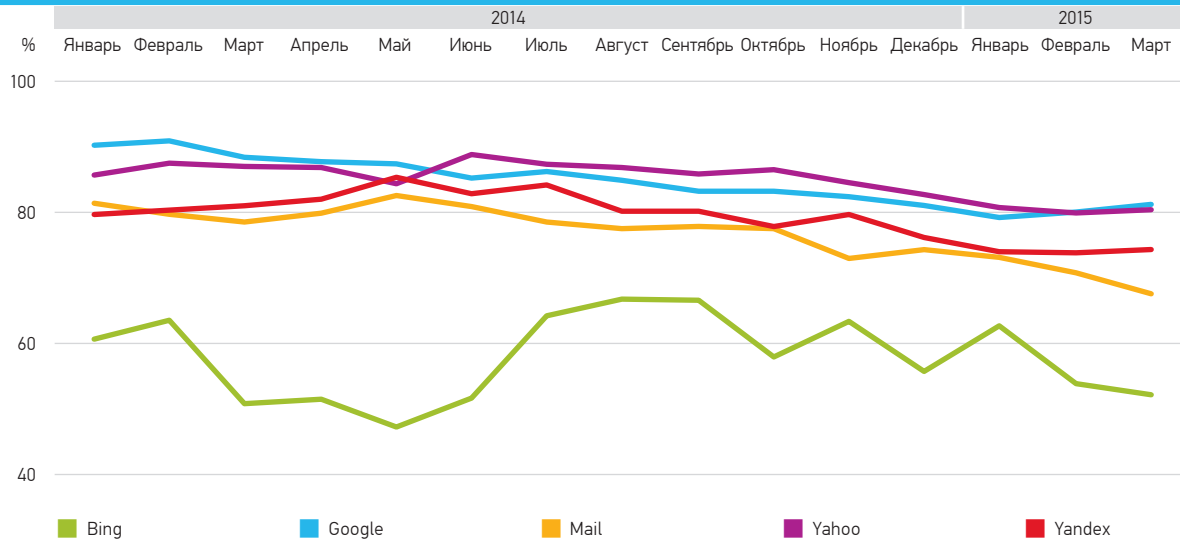
АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ПЕРИФЕРИЙНЫХ САЙТОВ

Периферийные сайты — это малоизвестные, небольшие сайты, чаще всего из регионов России. Очевидно, поисковые машины находят такие сайты хуже, чем крупные и популярные.

Если в 2013 году поисковики небольшими темпами улучшали результаты поиска по периферийным сайтам, а Яндексу удавалось достичь результата за 90%, то в 2014 году тенденция приобрела обратный характер. К весне 2015 результаты всех поисковиков находятся около уровня в 80% (рис. 2).

Это снижение может быть связано со все растущей ролью поведенческих факторов (в том числе кликов пользователей на странице результатов поиска) в поисковом ранжировании, что помогает попадать в выдачу популярным посещаемым сайтам, но отрицательно сказывается на позициях малопопулярных сайтов.

Рисунок 3. Качество внутрисайтовой навигации

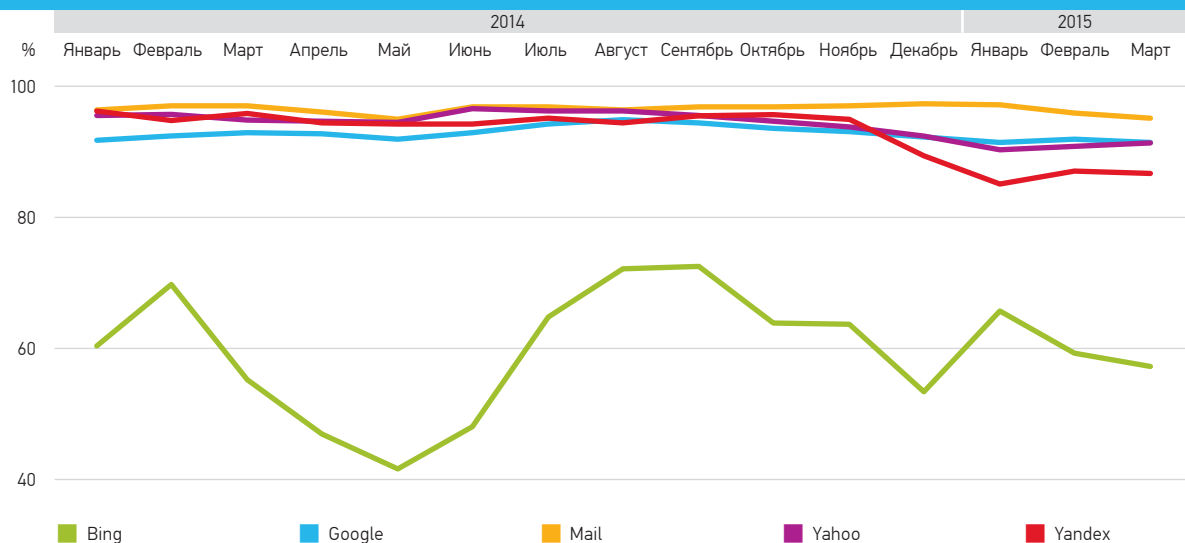


АНАЛИЗАТОР КАЧЕСТВА ВНУТРИСАЙТОВОЙ НАВИГАЦИИ

Анализатор качества внутрисайтовой навигации оценивает, насколько успешно поисковики понимают запросы, направленные на поиск определенной страницы (раздела) сайта. Например, по запросу «мегафон отправить смс» релевантной будет та страница на сайте Мегафона, с которой можно отправить смс, а главная (или любая другая) страница сайта почти бесполезна.

Лучше всех ссылки на разделы сайтов находит Google, Яндекс и Mail.ru немного отстают (рис. 3). Но к концу 2014 года показатели всех поисковиков стали ухудшаться. Отчасти это можно связать с тем, что некоторые маркеры в данном анализаторе устарели.

Рисунок 4. Качество поиска персон

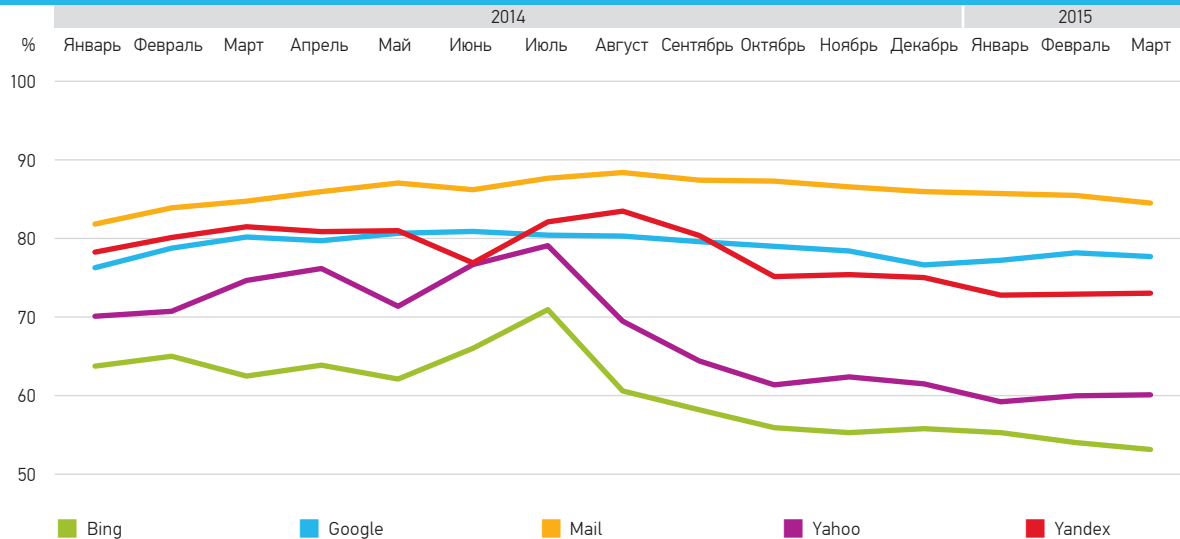


12

АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ПЕРСОН

Качество поиска персональных страниц у популярных в Рунете поисковиков достаточно высоко, хотя и не дотягивает до 100%. Отметим, что в 2014 году в лидеры по этому показателю вышел поиск Mail.ru (рис. 4).

Рисунок 5. Качество поиска персональных блогов

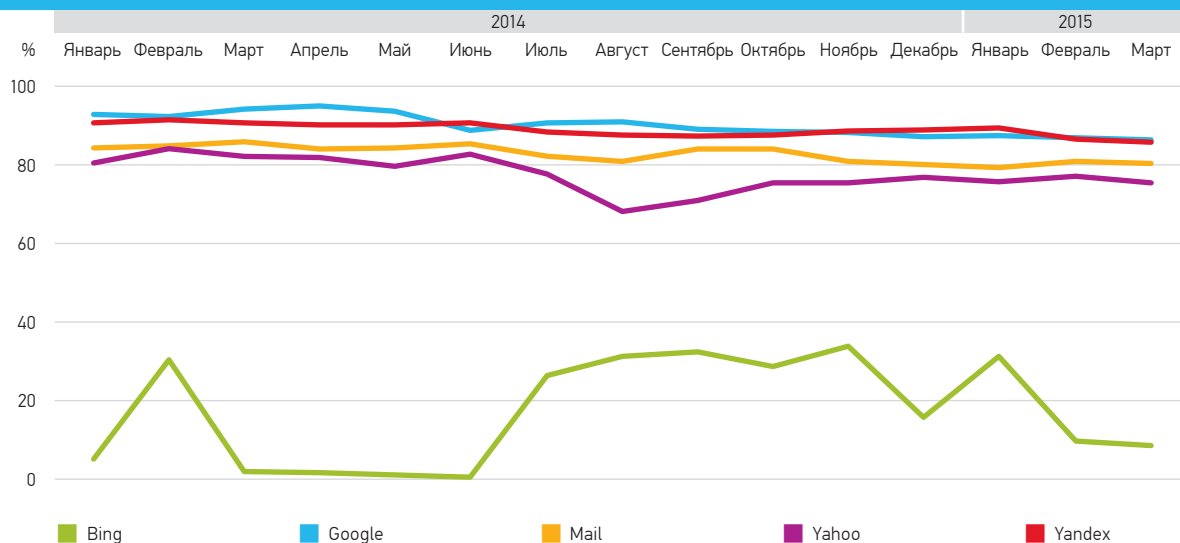


АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ПЕРСОНАЛЬНЫХ БЛОГОВ

Анализатор качества поиска персональных блогов аналогичен анализатору поиска персон, но его запросы — это люди, представленные в Рунете в первую очередь своими блогами. Соответственно, и релевантный результат — не персональная страница, а этот блог.

Проделанная разработчиками и программистами работа помогла Mail.ru занять в конце 2013 года первое место по этому показателю и удерживать его в течение всего 2014 года (рис. 5).

Рисунок 6. Качество навигационного поиска в социальных сетях



14

АНАЛИЗАТОР КАЧЕСТВА НАВИГАЦИОННОГО ПОИСКА В СОЦИАЛЬНЫХ СЕТЯХ

Этот анализатор похож на базовый анализатор качества навигационного поиска, но оценивает умение поисковиков находить не сайт, а страницу организации (или бизнеса) в той или иной социальной сети, когда эта страница заменяет официальный сайт.

Анализатор был запущен в августе 2013 года и сразу показал, как за 4 месяца команда Mail.ru заметно улучшила этот аспект работы своего поисковика. В результате, в 2014 году поиск Mail.ru стабильно занимал третье место по этому показателю, уступая лидерам несколько процентных пунктов. (рис. 6).

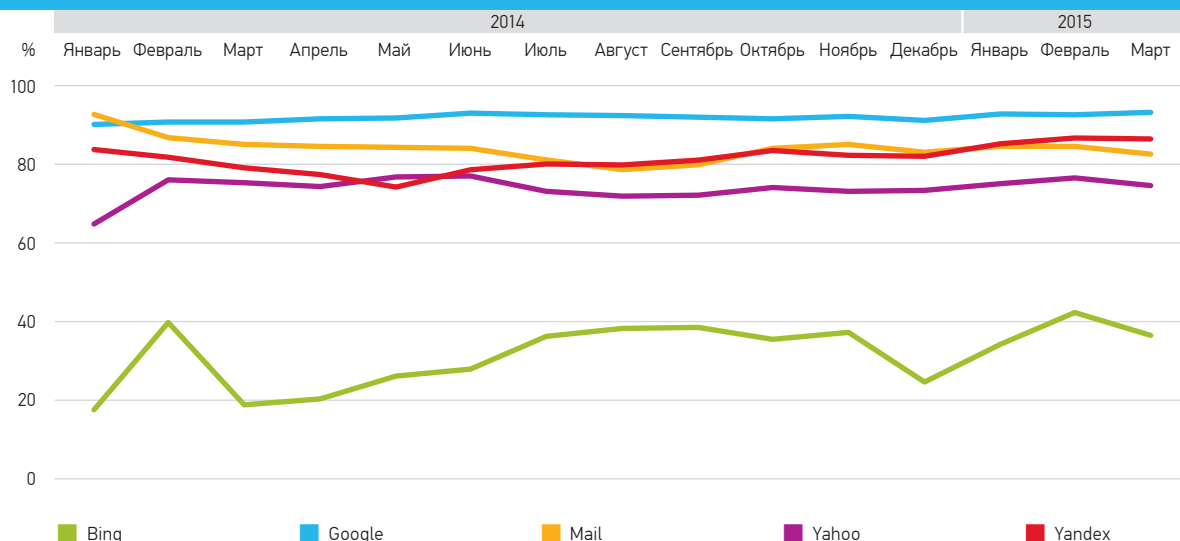
3.

КАЧЕСТВО ПОИСКА ПО ИНФОРМАЦИОН- НЫМ ЗАПРОСАМ

ИНФОРМАЦИОННЫМИ ЗАПРОСАМИ НАЗЫВАЮТ ТЕ, С ПОМОЩЬЮ КОТОРЫХ ПОЛЬЗОВАТЕЛЬ ПЫТАЕТСЯ НАЙТИ КАКУЮ-ТО ИНФОРМАЦИЮ: ОТВЕТ НА ВОПРОС, РЕШЕНИЕ ПРОБЛЕМЫ, КАКОЙ-НИБУДЬ ТЕКСТ И Т. П. УЖЕ ИЗ ЭТОГО ОПРЕДЕЛЕНИЯ ВИДНО, ЧТО ПОДОБНЫЕ ЗАПРОСЫ БЫВАЮТ ОЧЕНЬ РАЗНЫМИ, А ИХ ФОРМУЛИРОВКИ ДОВОЛЬНО РАСПЛЫВЧАТЫ. СРЕДИ ПОИСКОВЫХ ЗАПРОСОВ ИНФОРМАЦИОННЫЕ ВСТРЕЧАЮТСЯ ЧАЩЕ ВСЕГО.

Поскольку информационный поиск очень разнообразен, оценить его качество достаточно сложно. В большой степени эту задачу решает ассессорский анализатор: как и в поиске в целом, информационные запросы в нем преобладают. Но оценить отдельные аспекты информационного поиска можно, и это делают анализаторы соответствующей группы.

Рисунок 7. Анализатор качества поиска цитат

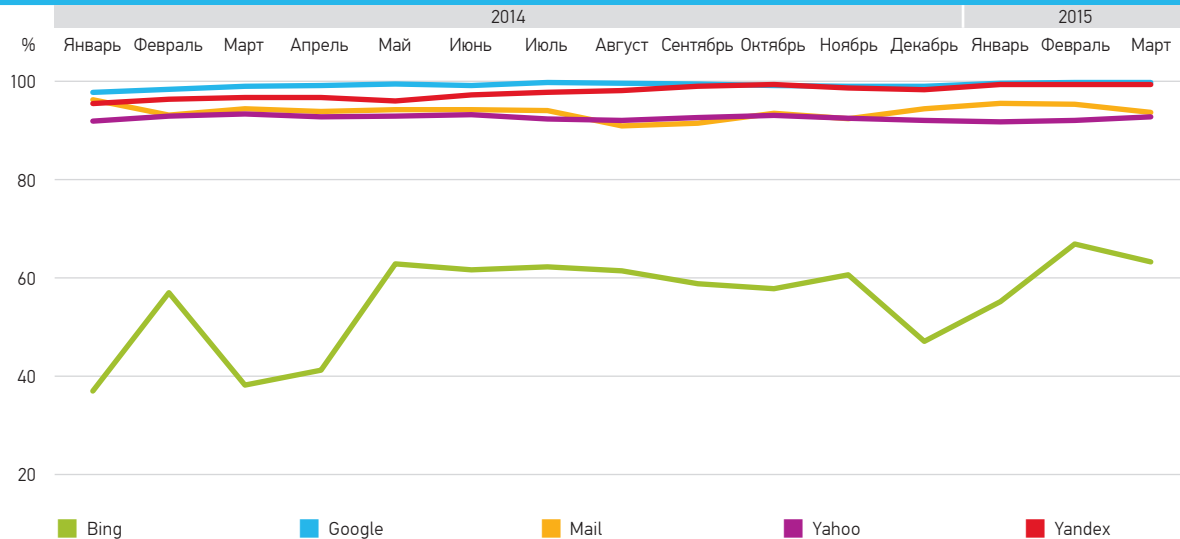


АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ЦИТАТ

Один из самых простых для оценивания аспектов информационного поиска — цитатный поиск. Если поисковый запрос является точной цитатой из некоторого произведения, то по нему обязательно должен находиться если не полный текст цитируемого произведения, то хотя бы название и автор. На сайте analyzethis.ru на дополнительных вкладках анализатора качества поиска цитат позволяет увидеть, насколько часто поисковики находят то и другое по отдельности, а суммарный результат анализатора — доля найденных страниц, где присутствует хотя бы что-нибудь одно.

В 2013 году только у Mail.ru оценка качества цитатного поиска превышала 90%. Но с февраля 2014 года на первое место вышел Google (рис. 7), в то время как результаты других поисковиков стали заметно хуже. Яндекс после провала в мае стал улучшать качество поиска цитат, и в 2015 году будет интересно наблюдать, удастся ли ему догнать лидера.

Рисунок 8. Анализатор качества поиска крылатых слов

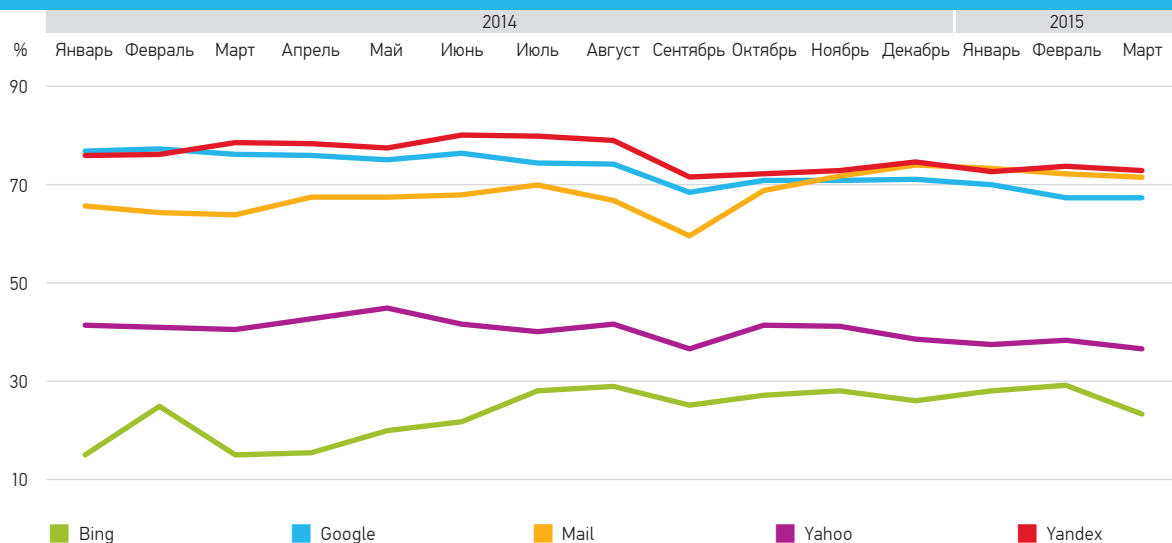


АНАЛИЗАТОР КАЧЕСТВА ПОИСКА КРЫЛАТЫХ СЛОВ

Очевидно, что вводя в поисковую строку какую-то крылатую фразу, пользователь хочет узнать толкование фразы или ее происхождение (автора и название произведения, если это цитата из литературного произведения), а не найти тексты, где она используется. Анализатор качества поиска крылатых фраз оценивает поисковики по тому, какая доля найденных по таким запросам страниц содержит указание на происхождение фразы (засчитываются как автор и название, так и исходный текст произведения).

Если в 2013 году в Яндексе, Google и Mail.ru были примерно одинаковые показатели в качестве поиска крылатых выражений — 95%, то в 2014 году лидеры (Яндекс и Google) улучшили результаты почти до 100%, а Yahoo и Mail.ru остались на прежнем уровне (рис. 8).

Рисунок 9. Анализатор качества поиска ответов



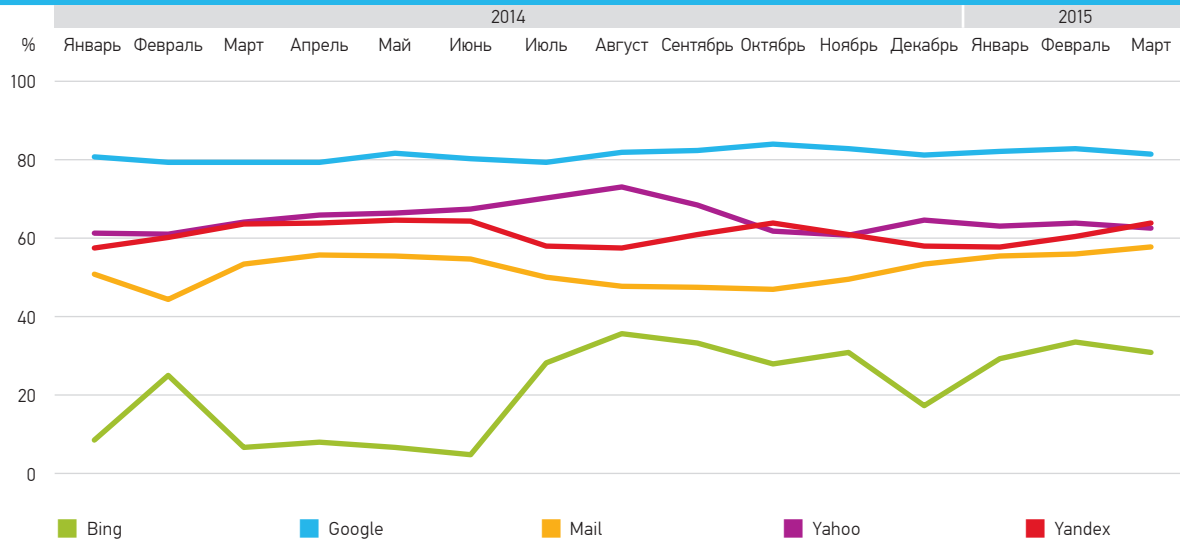
18

АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ОТВЕТОВ

Для этого анализатора подобраны запросы, подразумевающие явный ответ сразу на странице результатов поиска. Поэтому результат анализатора зависит от того, насколько высоко расположен снипет (цитата с найденного сайта) с правильным ответом («хелперы» засчитываются как первая позиция). На сайте analyzethis.ru можно посмотреть дополнительные вкладки этого анализатора, которые проверяют, в том числе и наличие ответов на страницах, найденных поисковиком.

К сожалению, заметных улучшений поиска ответов в 2014 году почти не было, единственное исключение — Mail.ru, которому удалось даже выйти на первое место по этому показателю. Впрочем, большую роль тут сыграло падение результатов лидеров (рис. 9).

Рисунок 10. Анализатор качества поиска оригиналов новостей



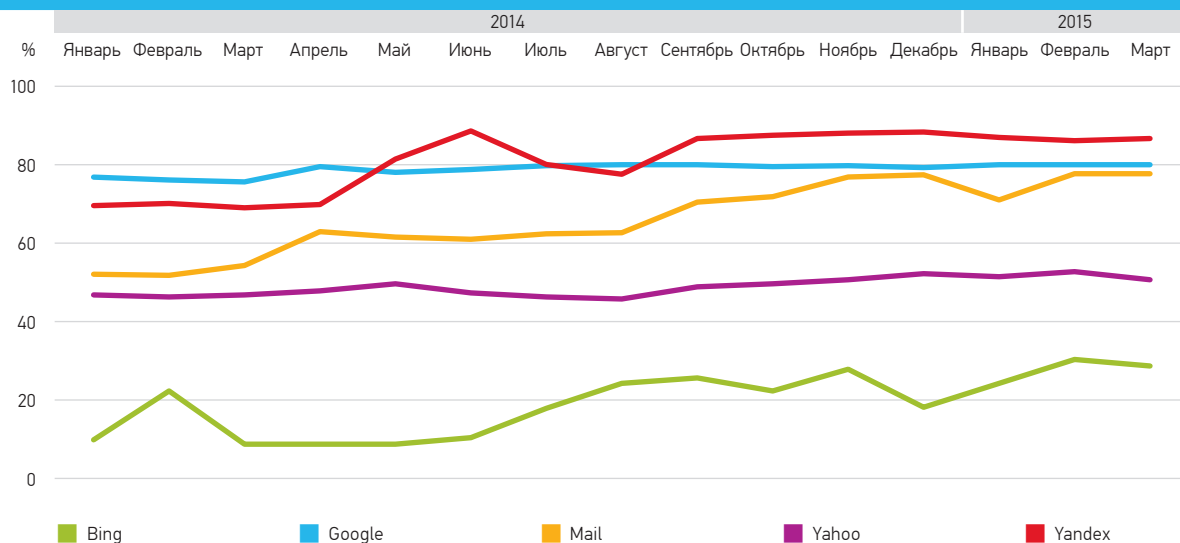
АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ОРИГИНАЛОВ НОВОСТЕЙ

Редкий владелец сайта будет обрадован, когда оригинальный текст, опубликованный им на своем сайте, оказывается без разрешения опубликован где-то еще. Еще хуже, если при этом поисковые системы показывают этот украденный текст в результатах поиска выше оригинала. К сожалению, это очень распространенное явление.

Запросы созданного нами анализатора — цитаты из новостных текстов, сайт-первоисточник которых достоверно известен. Анализатор ежедневно проверяет позиции этих сайтов в выдаче и оценивает долю запросов, по которым оригиналы находятся выше копий.

По сравнению с 2013 годом ситуация принципиально не изменилась: Google по-прежнему справляется с поиском оригиналов заметно лучше конкурентов (рис. 10). Интересно, что борьба за второе место идет между Яндексом и Yahoo.

Рисунок 11. Анализатор качества поиска адресов



20

АНАЛИЗАТОР КАЧЕСТВА ПОИСКА АДРЕСОВ

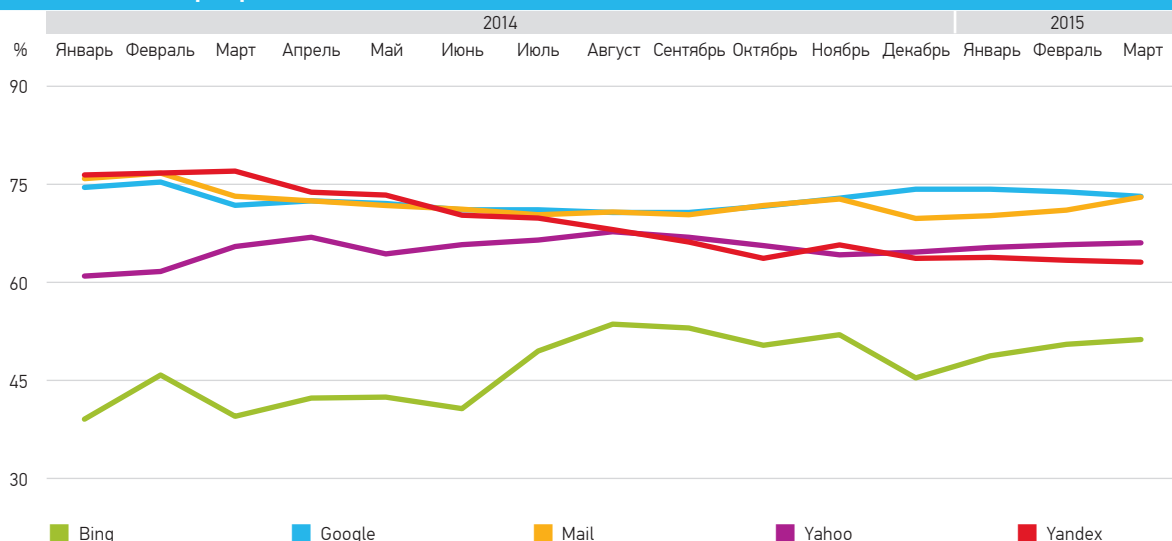
В 2013 году был запущен пятый анализатор информационного поиска, оценивающий качество нахождения адреса организации по ее названию и географическому ориентиру. Анализатор оценивает, насколько высоко поисковики показывают нужный адрес на странице выдачи, т.е. в хелперах и снипетах.

По сравнению с предыдущим годом результаты поисковиков стали немного лучше. Так, скачкообразные улучшения в Яндексе помогли ему преодолеть уровень в 80% и вывели его на первое место (рис. 11). В то же время Mail.ru показывал стабильный рост и смог почти догнать Google.

4.

КАЧЕСТВО ПОИСКА ПО ТРАНЗАКЦИОН- НЫМ ЗАПРОСАМ

СУЩЕСТВУЕТ МНОГО ВАРИАНТОВ ТРАНЗАКЦИОННЫХ ЗАПРОСОВ. ЭТО МОЖЕТ БЫТЬ ПОИСК ПЕСНИ ИЛИ ФИЛЬМА, А МОЖЕТ — ДОСТАВКИ ЦВЕТОВ. ВСЕ ЭТИ ЗАПРОСЫ ОБЪЕДИНЯЕТ ТО, ЧТО ОНИ ПОДРАЗУМЕВАЮТ КАКОЕ-ЛИБО ДЕЙСТВИЕ: «СКАЧАТЬ», «ПОСЛУШАТЬ», «КУПИТЬ» И ПР. ПЕРВЫЙ АНАЛИЗАТОР КАЧЕСТВА ТРАНЗАКЦИОННОГО ПОИСКА ПОЯВИЛСЯ В КОНЦЕ 2012 ГОДА, А В НАЧАЛЕ 2013 БЫЛ РАЗДЕЛЕН НА ДВА РАЗНЫХ.

Рисунок 12. Анализатор качества поиска официальных версий программного обеспечения


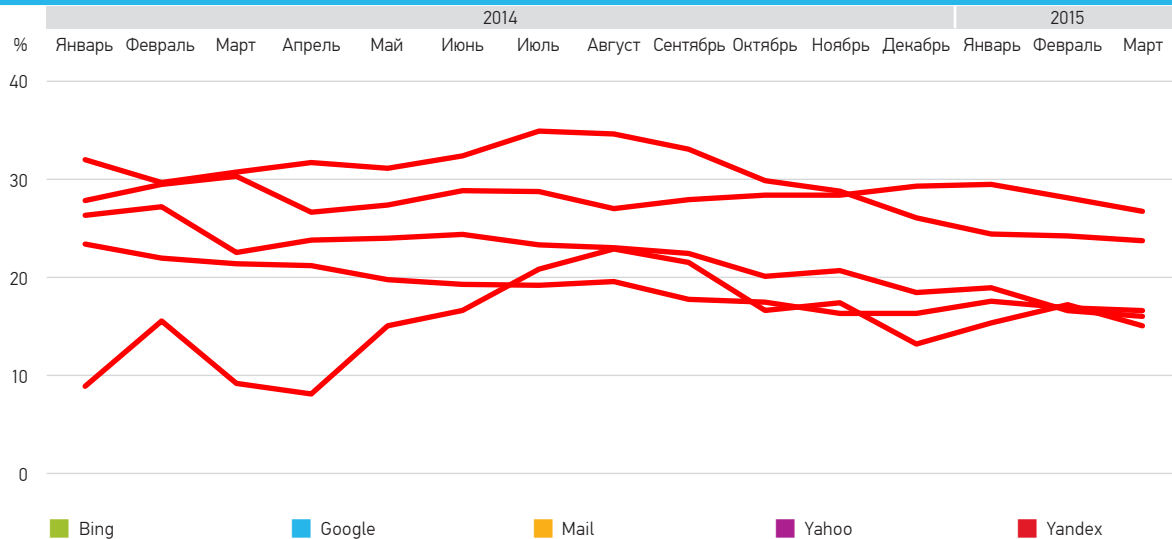
22

АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ОФИЦИАЛЬНЫХ ВЕРСИЙ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Анализатор качества поиска официальных версий ПО оценивает способность поисковиков находить ссылки на официальные сайты разработчиков, где популярные программы можно скачать безопасно и легально. Основная задача поисковиков — поставить в результатах поиска эти сайты выше сайтов-паразитов, собирающих трафик по таким запросам. На подобных сайтах пользователю будут не только навязчиво показывать порой очень раздражающую рекламу, но и могут при скачивании заразить его компьютер различными вирусами.

Если в 2013 году все поисковики постепенно улучшали поиск по запросам, нацеленным на скачивание программ, то в 2014 году по данным анализатора результаты скорее ухудшались. Стоит отметить падение Яндекса на предпоследнее место (рис. 12). Любопытно, что хотя поисковики внимательно следят за тем, чтобы в результатах поиска было как можно меньше опасных или насыщенных рекламой сайтов (подробнее об этом ниже), верхние позиции в результатах поиска по запросам, связанным со скачиванием программ, нередко достаются сайтам-посредникам.

Рисунок 13. Анализатор качества поиска официальных версий контента



АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ОФИЦИАЛЬНЫХ ВЕРСИЙ КОНТЕНТА

Аналогичные проблемы возникают и при поиске контента: способность находить среди горы мусора и сомнительных сайтов ссылки на официальные (или хотя бы легальные) источники аудио и видео контента и ставить их на первые места в выдаче является важным умением поисковика.

С нахождением легального контента поисковики справляются еще хуже: ни у одного из них показатель не превышает 35% (рис. 13). Очевидно, что проблема тысяч сайтов-пустышек, мешающих высоко ранжировать официальные источники контента, о которой мы писали год назад, никуда не исчезла.

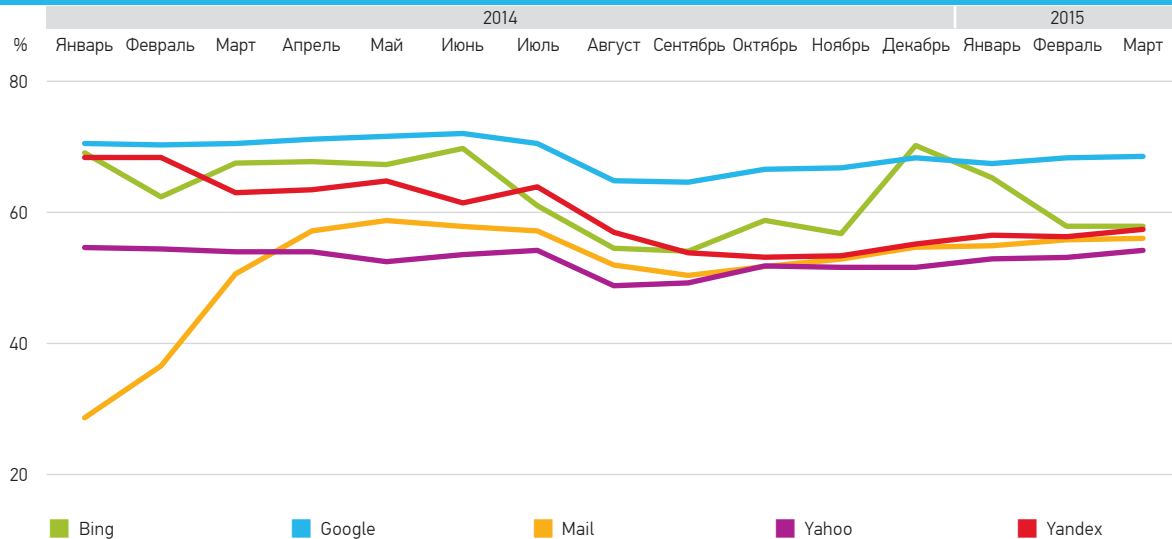
5.

АКТУАЛЬНОСТЬ РЕЗУЛЬТАТОВ ПОИСКА

24

ОТВЕТ ПОИСКОВОЙ МАШИНЫ НА ЗАПРОС ДОЛЖЕН БЫТЬ НЕ ПРОСТО РЕЛЕВАНТНЫМ, НО И АКТУАЛЬНЫМ, СВЕЖИМ. ПОИСКОВЫЕ РОБОТЫ ДОЛЖНЫ ПОСТОЯННО СЛЕДИТЬ ЗА ИЗМЕНЕНИЯМИ В ИНТЕРНЕТЕ И ОБНОВЛЯТЬ ПОИСКОВЫЙ ИНДЕКС. СВЕЖИЕ РЕЗУЛЬТАТЫ ОСОБЕННО ВАЖНЫ, ЕСЛИ РЕЧЬ ИДЕТ ОБ ИЗМЕНЕНИЯХ, ПРОИЗОШЕДШИХ НЕ ТОЛЬКО НА ВЕБ-СТРАНИЦЕ, НО И В ЖИЗНИ.

Рисунок 14. Анализатор актуальности поиска: должности



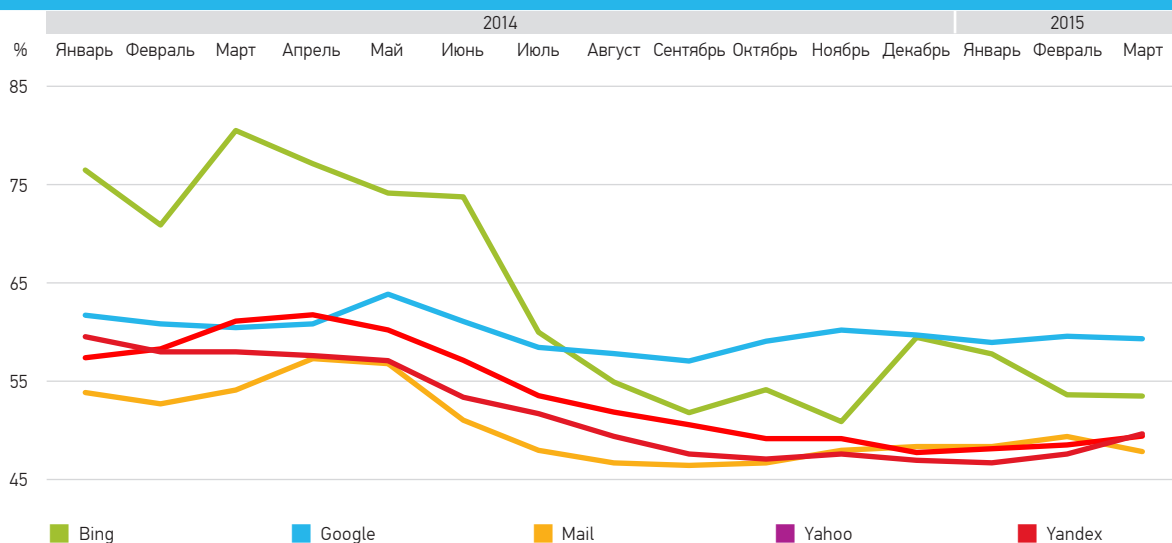
АНАЛИЗАТОР АКТУАЛЬНОСТИ ПОИСКА: ДОЛЖНОСТИ

Анализатор актуальности должностей позволяет оценить, насколько быстро поисковики обновляют свой индекс, отслеживая информацию о людях, занимающих те или иные должности, отставках и назначениях. Разумеется, чем больше времени прошло с момента назначения, тем больше поисковики будут показывать актуальных ответов, поэтому список запросов приходится регулярно обновлять.

Например, после обновления в декабре 2013 года показатели всех поисковиков ощутимо упали, а падение Mail.ru и вовсе составило более 40%. Восстановить показатели Mail.ru удалось только к маю 2014 года.

Очередное обновление списка запросов наши специалисты осуществили в августе 2014 года. В этот раз расстановка сил не изменилась, хотя показатели предсказуемо снизились у всех поисковых машин (рис. 14), после чего снова стали улучшаться.

Рисунок 15. Анализатор актуальности поиска: телефоны



26

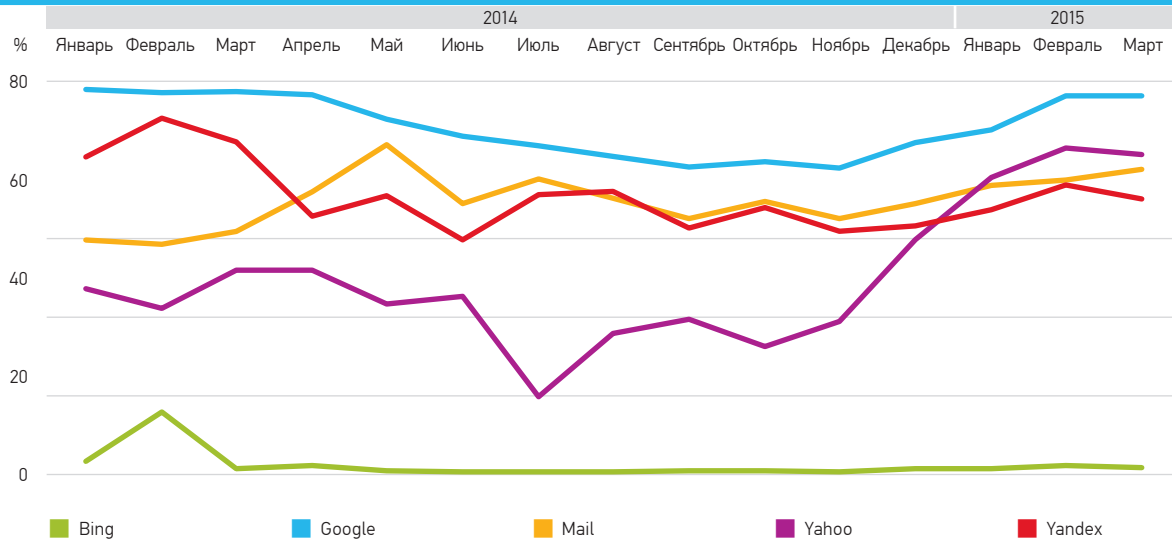
АНАЛИЗАТОР АКТУАЛЬНОСТИ ПОИСКА: ТЕЛЕФОНЫ

Анализатор актуальности результатов поиска телефонов также оценивает свежесть поисковой выдачи, но использует другие запросы: название организаций + нередко слово-уточнение «телефон».

Запросы этого анализатора тоже регулярно обновляются. В июне 2014 года на графике видны следы подобного обновления (рис. 14): результаты всех поисковиков

синхронно упали. Обращает на себя внимание некоторое отставание Яндекса в восстановлении показателей в январе—феврале 2015 года, и его заметное отставание от Google в этом анализаторе. В целом ситуация хуже чем год назад: в 2013 году результаты поисковиков были близки к 70%, а в 2014 году только Google достигал уровня в 60%, значения у остальных поисковиков ниже 50%.

Рисунок 16. Анализатор скорости индексации



АНАЛИЗАТОР СКОРОСТИ ИНДЕКСАЦИИ

Чтобы понять насколько быстро могут обновляться результаты поиска, нужно оценить скорость индексации роботами новых страниц с сайтов. Для этого был создан анализатор скорости индексации, который узнает из специальных файлов sitemap.xml о появлении свежих страниц на крупных, часто обновляемых сайтах.

Анализатор в течение 30 дней проверяет, находится ли данная страница в поиске. Интересно, что в течение 30 дней в индекс Google попадает в среднем только 80% новых страниц (рис. 15), а осенью 2014 года попадало еще меньше страниц. Скорость индексации Яндекса непостоянна, но стала заметно ниже, чем в начале 2014 года. В то же время поиск Mail.ru улучшил индексацию.

Этот анализатор дополнительно показывает, какая доля новых страниц попала в индекс в течение первых 2 дней (а так же в течение 3—6 дней и в первые две недели, эти результаты можно посмотреть на дополнительных вкладках на сайте analyzethis.ru). Отметим, что с такой быстрой индексацией новых документов лучше всех справляется Mail.ru, чей показатель достиг в 2014 году 70% и в начале 2015 продолжил рост. Яндекс и Google за короткий срок успевают проиндексировать в среднем 30—40% новых страниц.

6.

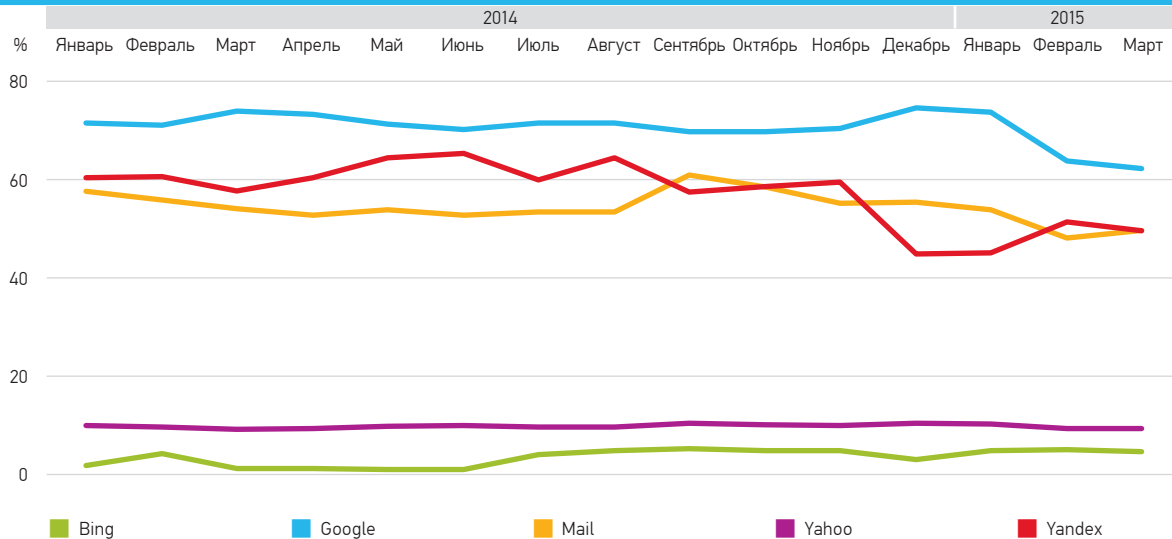
ПОЛНОТА И РАЗНООБРАЗИЕ РЕЗУЛЬТАТОВ ПОИСКА

28

ПО ДАННЫМ РАЗРАБОТЧИКОВ ПОИСКОВЫХ МАШИН, СРЕДНЯЯ ДЛИНА ПОИСКОВОГО ЗАПРОСА РАСТЕТ: В ПЕРВУЮ ОЧЕРЕДЬ ЭТО ОЗНАЧАЕТ, ЧТО ДОЛЯ ЗАПРОСОВ, СОСТОЯЩИХ ИЗ НЕСКОЛЬКИХ СЛОВ БОЛЬШЕ, ЧЕМ ОДНОСЛОВНЫХ. НАПРИМЕР, В ЯНДЕКСЕ СРЕДНЯЯ ДЛИНА ЗАПРОСА — 3,4 СЛОВА ([HTTP://COMPANY.YANDEX.RU/RESEARCHES/REPORTS/2013/YA_SEARCH_INTERESTS_2013.XML](http://COMPANY.YANDEX.RU/RESEARCHES/REPORTS/2013/YA_SEARCH_INTERESTS_2013.XML)).

Это значит, что поисковик должен справляться с самыми разными запросами, а не только с наиболее частыми. Для этого нужно уметь широко трактовать запросы, когда их смысл не однозначен, и иметь большой охват (полноту индекса), чтобы находить ответ даже на самые специфические запросы.

Рисунок 17. Анализатор полноты индекса

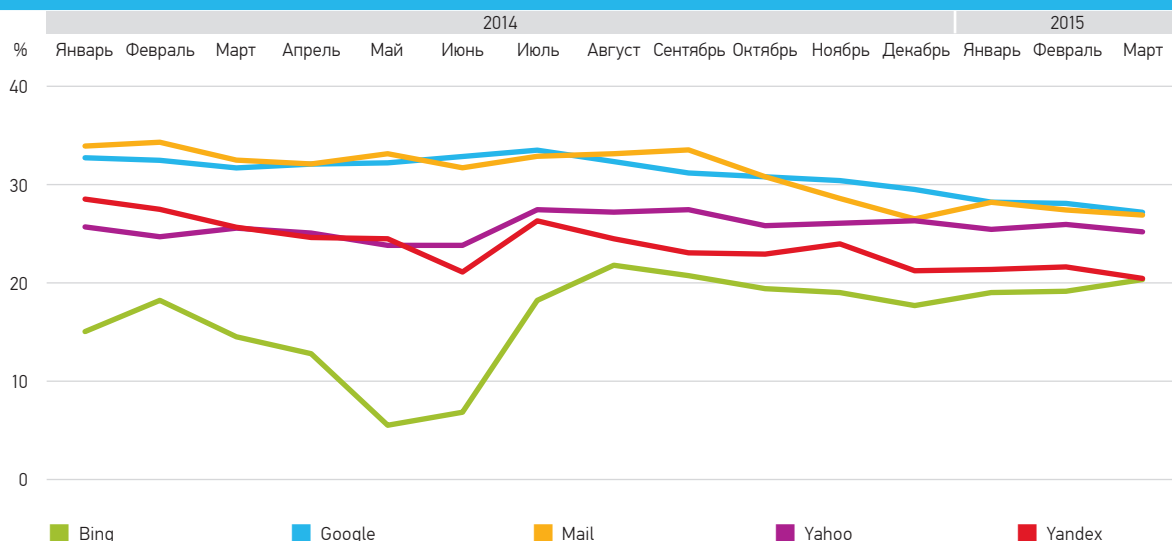


АНАЛИЗАТОР ПОЛНОТЫ ИНДЕКСА

Размер индекса — показатель количества сайтов, известных поисковику. Чем он полнее, тем меньше запросов останется без ответа и тем точнее будет ответ на конкретизированные запросы. Для оценки полноты индекса мы подбираем специальные слова, которые редко встречаются как в поисковых запросах, так и на сайтах в Сети. Стоит отметить, что анализатор оценивает индекс поисковиков относительно друг друга, поскольку у нас нет способа узнать обо всех страницах в Интернете. Его значения можно считать верхней оценкой доли страниц Рунета, известных поисковику.

В 2013 году относительная полнота индекса Mail.ru достигла уровня Яндекса, что стало заметным событием в индустрии поиска. В 2014 году полнота поисковиков относительно друг друга почти не менялась, только в конце года полнота Яндекса заметно упала (рис. 17).

Рисунок 18. Анализатор качества тематического поиска



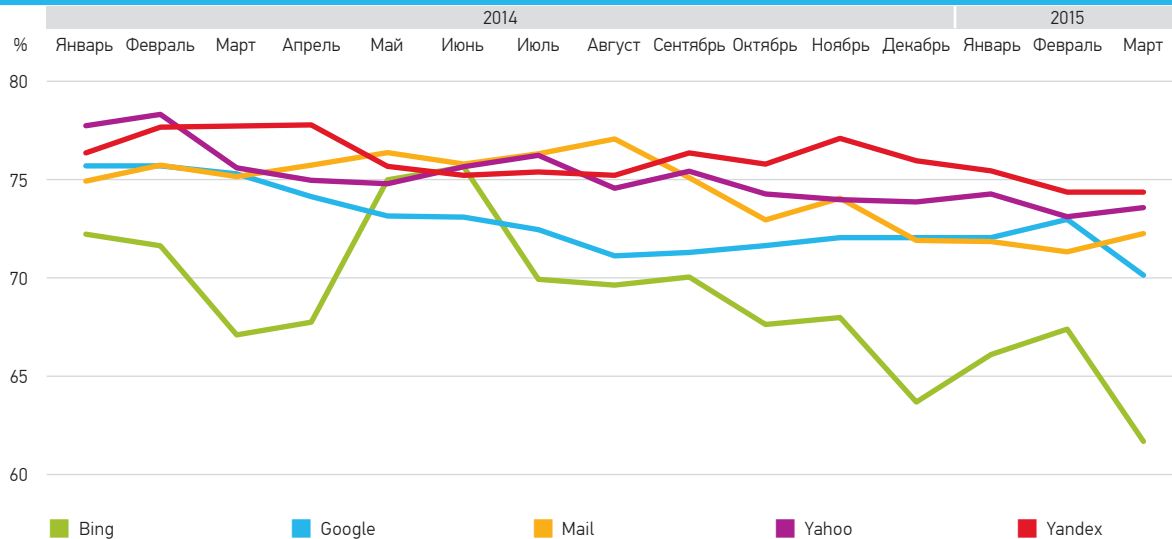
30

АНАЛИЗАТОР КАЧЕСТВА ТЕМАТИЧЕСКОГО ПОИСКА

Анализатор тематического поиска оценивает способность поисковых машин находить все сайты организаций, соответствующих конкретному запросу. Например, мы можем вручную составить список «Детских спортивных школ республики Татарстан» или «Пермских химических заводов», особенно, имеющих свои сайты, и сравнить его с тем, что выдают поисковики в ответ на эти запросы.

Ситуация с поиском сайтов организаций по тематическим запросам в 2014 году еще немного ухудшилась. У лидеров тематического поиска, Google и Mail.ru, результат к концу года опустился ниже 30% (рис. 18). Впрочем, это не означает, что результаты поиска по этим запросам действительно плохи, так как становится все больше сайтов-агрегаторов, которые занимают места в поисковой выдаче, т.к. помогают найти сразу все нужные страницы.

Рисунок 19. Анализатор качества поиска омонимов



АНАЛИЗАТОР КАЧЕСТВА ПОИСКА ОМОНИМОВ

Мы считаем, что по запросам, состоящим из одного слова, имеющего несколько смыслов, поисковики должны находить страницы или сайты, отвечающие каждому из возможных значений.

В 2014 году качество поиска по словам-омонимам немного ухудшилось: к концу года поисковики стали упускать около четверти возможных смыслов запро-

са (рис. 19). Интересно, что для запросов, имеющих 2 смысла, в результатах поиска почти всегда встретятся оба значения. А по запросам-омонимам с 3 и тем более 4 разными смыслами поисковики часто упускают отдельные возможные варианты значений слова.

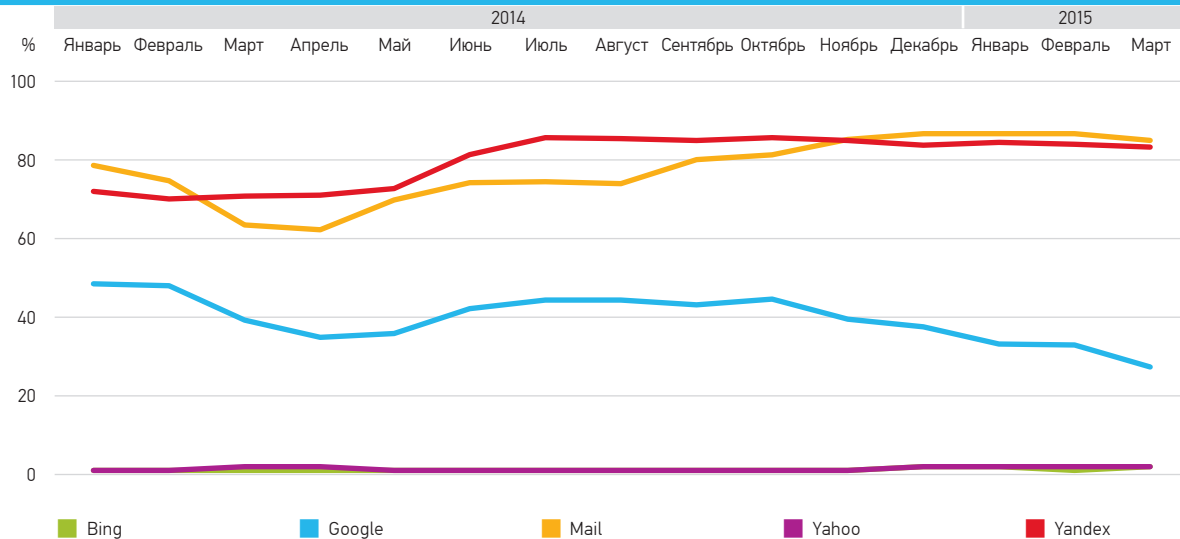
7.

ПОИСК ИЗ РЕГИОНОВ

32

В 2010 ГОДУ ЯНДЕКС ПРЕДСТАВИЛ ИССЛЕДОВАНИЕ РЕГИОНАЛЬНЫХ ОСОБЕННОСТЕЙ ПОИСКА, В КОТОРОМ ОТМЕЧАЛОСЬ, ЧТО ПРИМЕРНО 16% ЗАПРОСОВ К ПОИСКОВИКУ ЯВЛЯЮТСЯ РЕГИОНАЛЬНЫМИ — Т.Е. ВЫДАЧУ ПО НИМ НУЖНО СТРОИТЬ ИСХОДЯ ИЗ ТОГО, ИЗ КАКОГО ГОРОДА БЫЛ СДЕЛАН ЗАПРОС. ЭТО ОЧЕНЬ ЗАМЕТНАЯ ДОЛЯ ПОИСКОВОГО ТРАФИКА, КОТОРАЯ С ТЕХ ПОР ТОЛЬКО ПРОДОЛЖАЕТ РАСТИ. ПОЭТОМУ КАЧЕСТВО ПОИСКА В РЕГИОНАХ — ДОСТАТОЧНО ВАЖНЫЙ АСПЕКТ ОБЩЕГО КАЧЕСТВА ПОИСКА.

Рисунок 20. Анализатор качества регионального поиска



33

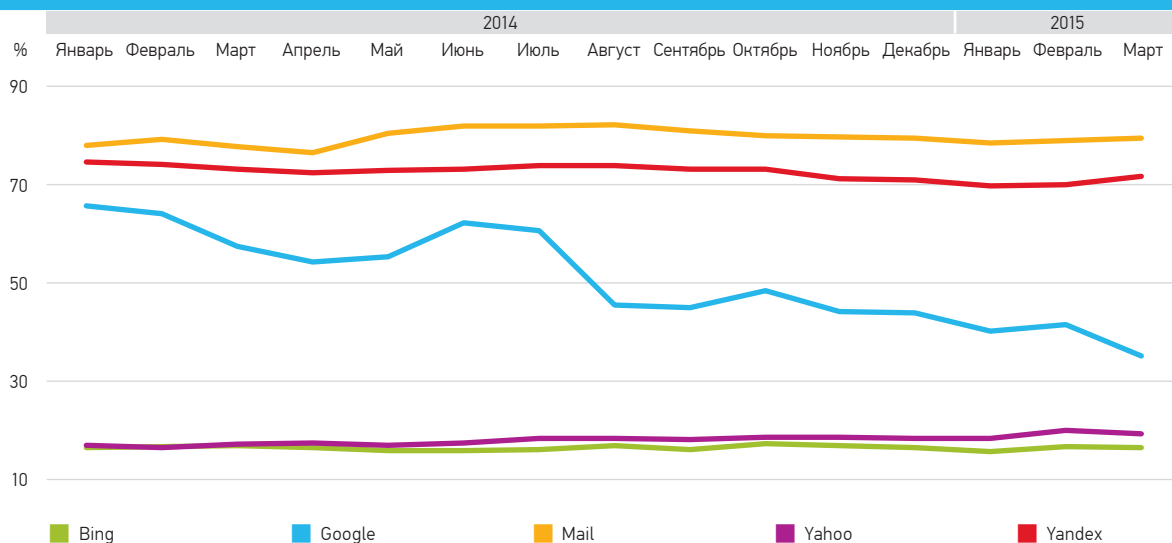
АНАЛИЗАТОР РЕГИОНАЛЬНОГО ПОИСКА

Анализатор регионального поиска подсчитывает долю региональных страниц в выдаче по таким информационным запросам, для которых географическое положение пользователя очевидно значимо (например, «адреса зоомагазинов» или «роллы доставка»).

Если в 2013 году Google удалось значительно улучшить этот аспект поиска, увеличив свой результат в анализаторе с 20% до 50%, то в 2014 году результаты снова

упали, до 30% (рис. 20). В то же время российские поисковики Яндекс и Mail.ru продолжали улучшать региональный поиск и довели этот показатель почти до 85%. Стоит отметить, что на сайте проекта анализаторов можно изучить результаты по каждому из 10 исследуемых городов.

Рисунок 21. Качество регионального навигационного поиска



34

АНАЛИЗАТОР КАЧЕСТВА РЕГИОНАЛЬНОГО НАВИГАЦИОННОГО ПОИСКА

Этот анализатор, как и общий анализатор навигационного поиска, проверяет нахождение сайта организации по ее названию. Отличие же в том, что должна найтись страница регионального отделения или региональной организации, т.е. своя для каждого из регионов. Например, по запросу «филармония» в Краснодаре и Самаре нужно показывать разные сайты.

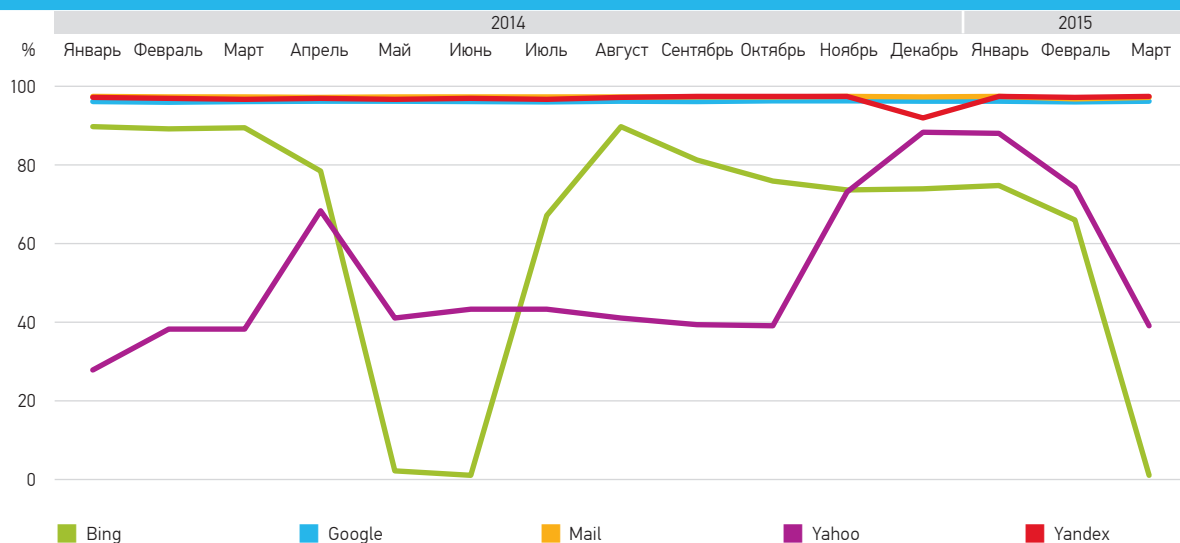
Качество региональной навигации у Яндекса и Mail.ru за год практически не изменилось, а вот результаты Google и здесь заметно снизились: с 65% до 35% (рис. 21). Синхронное падение поисковика в региональных анализаторах показывает, что это направление, скорее всего, перестало быть для Google интересным.

8.

ПОНИМАНИЕ ЗАПРОСА

КОНЕЧНО, ОТ ПОИСКОВЫХ СИСТЕМ ТРУДНО ТРЕБОВАТЬ ПОНИМАНИЯ СУТИ ФРАЗ И СМЫСЛА СЛОВ, НО ПОДСКАЗАТЬ, ИСПРАВИТЬ ОШИБКУ ТАМ, ГДЕ ЭТО НУЖНО, ИЛИ СВЯЗАТЬ МЕЖДУ СОБОЙ ПОХОЖИЕ ЗАПРОСЫ — ВСЕ ЭТО ПОИСКОВЫМ МАШИНАМ УЖЕ ПО СИЛАМ. В ЭТОЙ ГРУППЕ ОБЪЕДИНЕНЫ АНАЛИЗАТОРЫ, ИССЛЕДУЮЩИЕ ТАКИЕ УМЕНИЯ ПОИСКОВИКОВ.

Рисунок 22. Качество подсказок в поиске



36

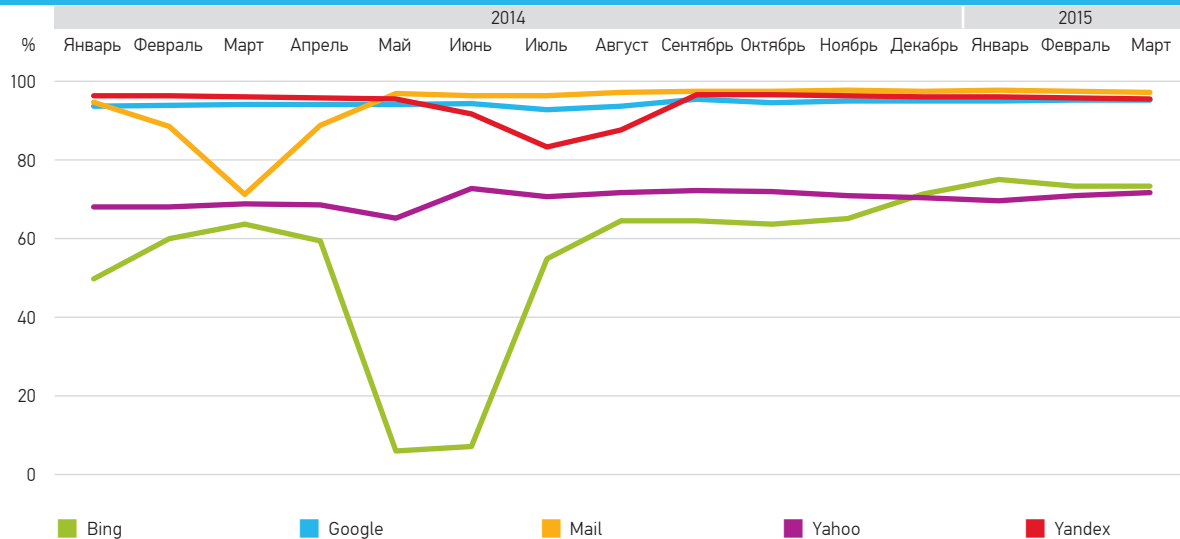
АНАЛИЗАТОР КАЧЕСТВА ПОДСКАЗОК

Когда пользователь вводит поисковый запрос с опечаткой, поисковик должен это понять или хотя бы предположить и либо сформировать выдачу с добавлением найденного по исправленному запросу, либо полностью заменить запрос на правильный.

Анализатор качества подсказок оценивает количество предлагаемых поисковином правильных подсказок (как в виде предположений, так и в виде замен запроса), если задать запрос с ошибкой.

Три лидирующих поисковика очень хорошо корректируют ошибки пользователей: в среднем незамеченными остаются не больше 3—5 опечаток из ста (рис. 22), причем этот уровень в Яндексе, Google и Mail.ru держится уже второй год.

Рисунок 23. Устойчивость к опечаткам

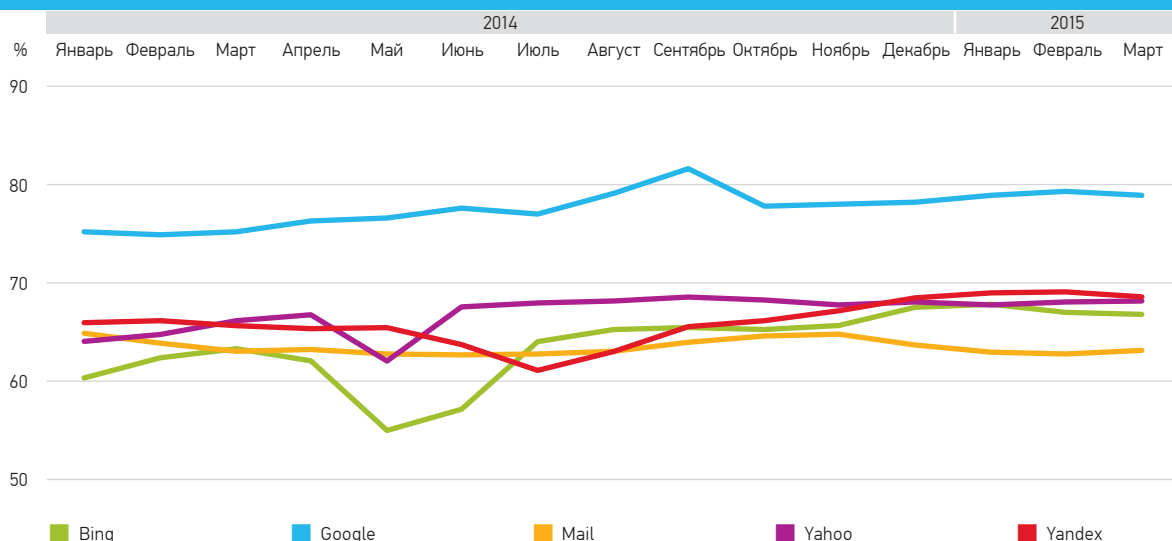


АНАЛИЗАТОР УСТОЙЧИВОСТИ К ОПЕЧАТКАМ В ЗАПРОСЕ

Если запрос с опечаткой становится полностью бессмысленным (например, «административный клдекс»), то очевидно, что выдача по нему должна совпадать с выдачей по аналогичному вопросу, заданному без ошибок. Анализатор сравнивает выдачу по ошибочному и по правильному запросам, считая долю совпадающих страниц (чем больше, тем лучше).

Лидеры (Яндекс, Google и Mail.ru) и здесь демонстрируют неплохие результаты: степень совпадения выдачи по запросам с ошибкой и без превышает 96%, т.е. расхождения будут встречаться в выдаче реже, чем 1 позиция на 2 запроса (рис. 23). Это еще немного реже, чем в конце 2103 г.

Рисунок 24. Уровень схожести выдачи по запросам-синонимам

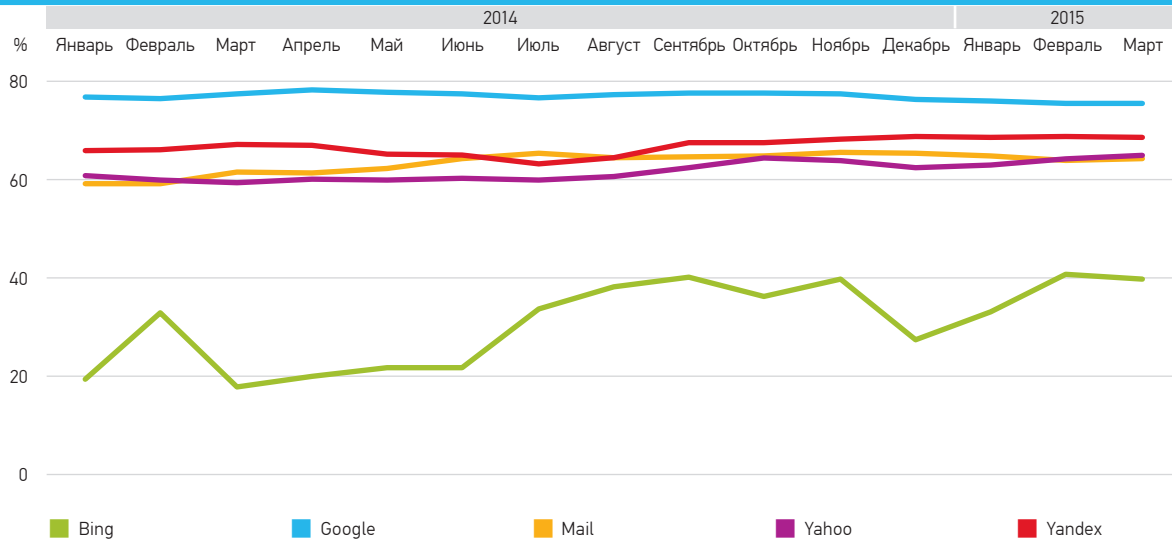


АНАЛИЗАТОР ЗАПРОСОВ-СИНОНИМОВ

Под запросами-синонимами мы понимаем в первую очередь не слова-синонимы, а поисковые запросы, которые очень похожи по смыслу, но чем-то немного различаются: перестановкой слов, использованием сокращений, транслитерацией, добавлением бессмысленных слов. В таких случаях естественно ожидать, что различия в выдаче тоже будут минимальными, например, по запросам «годовщины свадьбы» и «свадебные годовщины», «заказать жд билеты» и «заказ железнодорожных билетов», «манго одежда» и «одежда mango». Анализатор оценивает, насколько совпадают выдачи по таким синонимичным запросам.

На графике (рис. 24) видно, что в Google суть запроса играет самую большую роль: почти 80% результатов поиска по таким запросам совпадают. Значения Яндекса (69%) и Mail.ru (63%) показывают, что алгоритмы этих поисковиков делают более заметный акцент на сам текст запроса. Впрочем, доля различающихся страниц в выдаче Яндекса по похожим запросам уменьшалась в течение всего 2014 года.

Рисунок 25. Качество поиска по запросам с числами



АНАЛИЗАТОР ПОИСКА ПО ЗАПРОСАМ С ЧИСЛАМИ

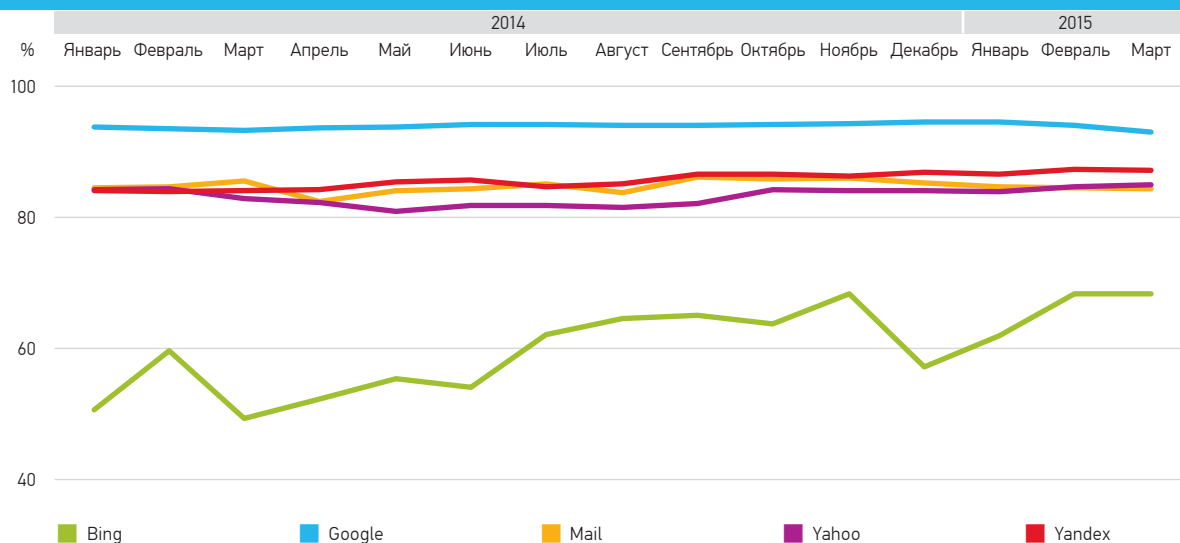
Очевидно, что поисковые машины должны обрабатывать запросы с числами с повышенным вниманием. Ведь запросы «поезд 017А» и «поезд 027А» — это принципиально разные запросы, а в запросе «новопассит 200 мл цена» число 200 не должно путаться с ценой.

В сентябре 2013 года было добавлено сразу два анализатора, отражающие умения поисковиков в этой области: анализатор поиска по запросам с числами и анализатор поиска моделей.

Анализатор поиска по запросам с числами проверяет каждую страницу из результатов поиска, чтобы понять контекст вхождения числа из запроса в текст страницы. Чем больше доля страниц, где число найдено «по делу», тем лучше.

Лидером по правильному поиску русскоязычных запросов с числами является Google, но даже у него более 20% найденных страниц содержат числа из запроса только в неправильном контексте, а иногда и просто случайно (рис. 25). Яндекс и Mail.ru постепенно улучшают этот аспект качества поиска.

Рисунок 26. Качество поиска моделей



40

АНАЛИЗАТОР ПОИСКА МОДЕЛЕЙ

Названия моделей — достаточно устойчивые сочетания, особенно в интернете, именно поэтому этот, по сути, частный случай поиска по запросам с числами был выделен в отдельный анализатор. И действительно, с поиском конкретных моделей техники или товаров поисковики ожидаемо справляются лучше, чем по запросам с числами в целом.

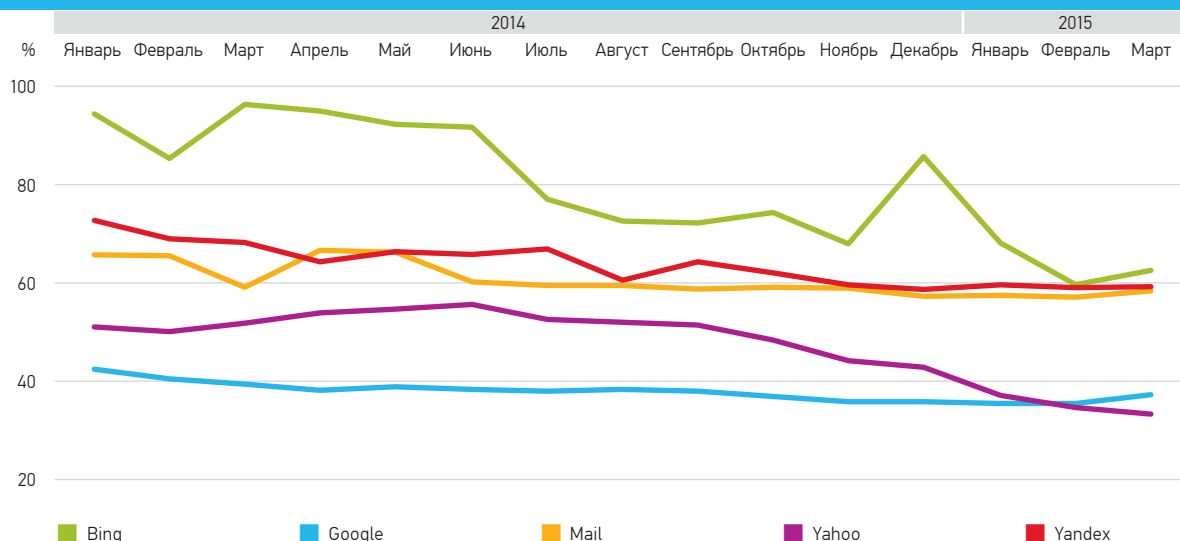
Google лидирует и тут, находя нужные модели на почти 95% страниц, показанных по соответствующим запросам (рис. 26). За 2014 год Яндексу удалось немного улучшить поиск моделей и приблизиться к уровню в 90%.

9.

АНАЛИЗАТОРЫ ОШИБОК

ОШИБКИ ПОИСКОВЫХ СИСТЕМ — ЭТО В ПЕРВУЮ ОЧЕРЕДЬ ОБРАТНАЯ СТОРОНА РАБОТЫ ПО ИНТЕРПРЕТАЦИИ САМЫХ РАЗНООБРАЗНЫХ ЗАПРОСОВ ПОЛЬЗОВАТЕЛЕЙ. ТАКИЕ ОШИБКИ МОГУТ БЫТЬ ДОСТАТОЧНО ЗАБАВНЫМИ, ЕСЛИ ИХ ПРИЧИНА ОЧЕВИДНА, НО НЕРЕДКО МОГУТ И РАЗДРАЖАТЬ ПОЛЬЗОВАТЕЛЯ, ОСОБЕННО, ЕСЛИ ИЗ-ЗА НИХ ЕМУ НЕ УДАЕТСЯ НАЙТИ НУЖНУЮ ИНФОРМАЦИЮ.

Рисунок 27. Количество ошибочных замен редких слов



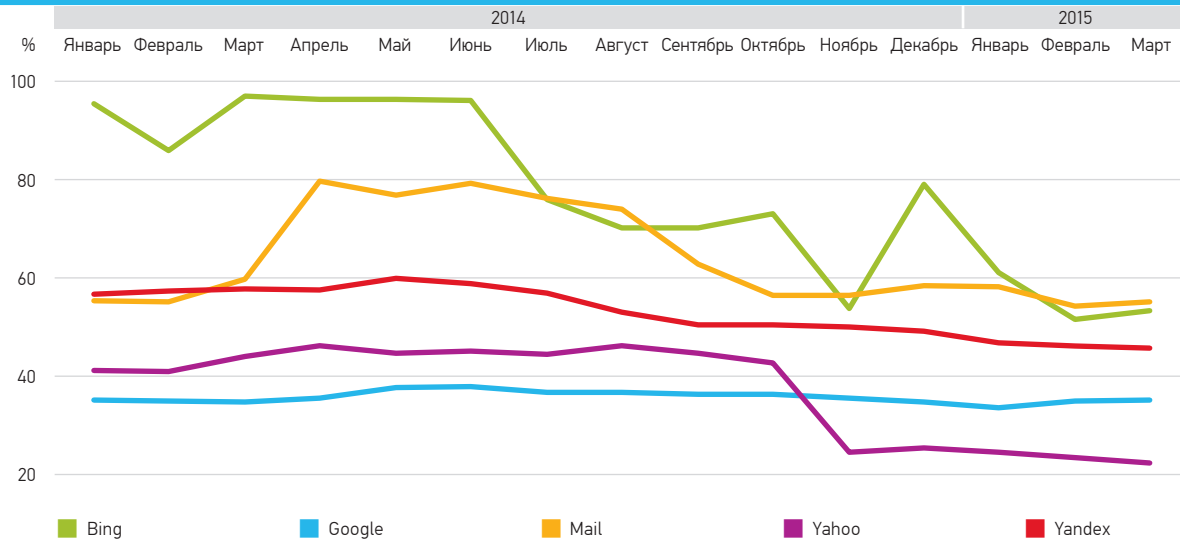
42

АНАЛИЗАТОР ЗАМЕНЫ РЕДКИХ СЛОВ

Создавая анализатор редких запросов, мы отобрали малоизвестные, но существующие, в том числе и в Интернете, слова, и оцениваем, насколько часто поисковики неоправданно заменяют их похожими, но более частотными. Анализатор проверяет наличие искомого слова в показанных поисковиками снипетах и вычисляет долю тех, где нужное слово не встретилось. Дополнительно вычисляется количество случаев, где поисковик прямо заменил или предложил заменить корректный с точки зрения потенциального пользователя запрос.

Несмотря на то, что Яндекс и в меньшей степени Mail.ru за год улучшили этот аспект качества поиска (доля снипетов без искомого слова снизилась до 60%), назвать результаты удовлетворительными все еще достаточно сложно (рис. 27). Даже у занимающего здесь первое место Google около 35% сайтов, найденных по запросам с редкими словами, не имеют к ним никакого отношения.

Рисунок 28. Количество ошибочных замен редких фраз

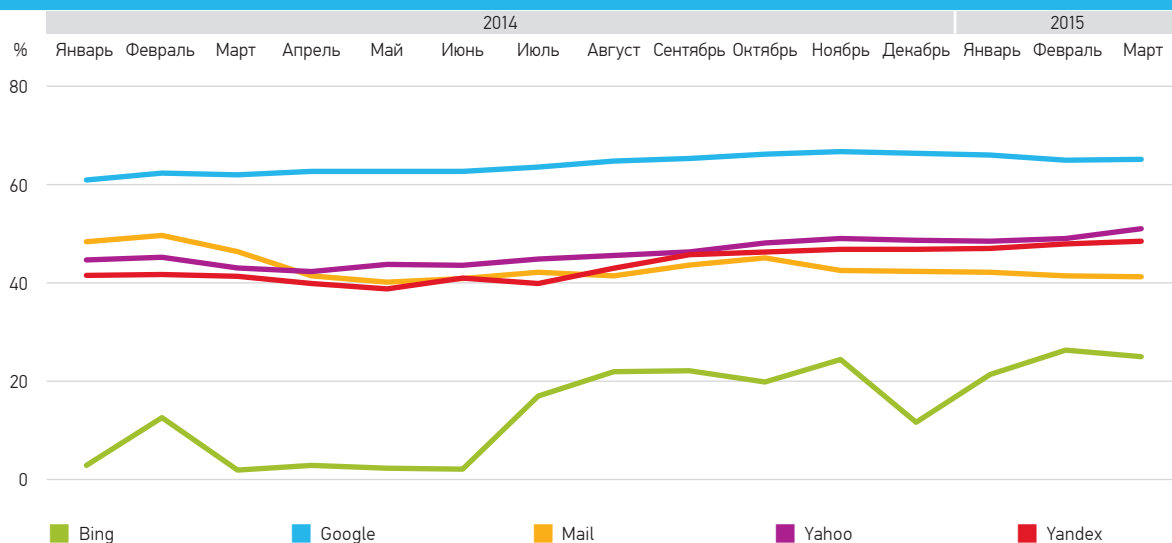


АНАЛИЗАТОР ЗАМЕНЫ РЕДКИХ ФРАЗ

Поисковики часто заменяют слова, которые кажутся им слишком редкими, даже в многословных запросах, хотя контекст зачастую полностью отрицает возможность такой замены, а ее ошибочность обычно видна невооруженным взглядом (ср. «озеро банук», заменяемое на «озеро банк»). Запрос в этом случае полностью теряет смысл, что может довольно сильно раздражать пользователей.

В целом ситуация здесь такая же, как и с ошибками в редких словах: сильнейшие поисковики делают довольно много неверных замен, от 35% у Google до 55% у Mail.ru (рис. 28). К концу 2014 года на первое место по этому показателю вышел Yahoo, но из-за меньшей полноты русскоязычного индекса он тоже часто не находит достаточного количества релевантных результатов.

Рисунок 29. Качество поиска неразрывных словосочетаний



44

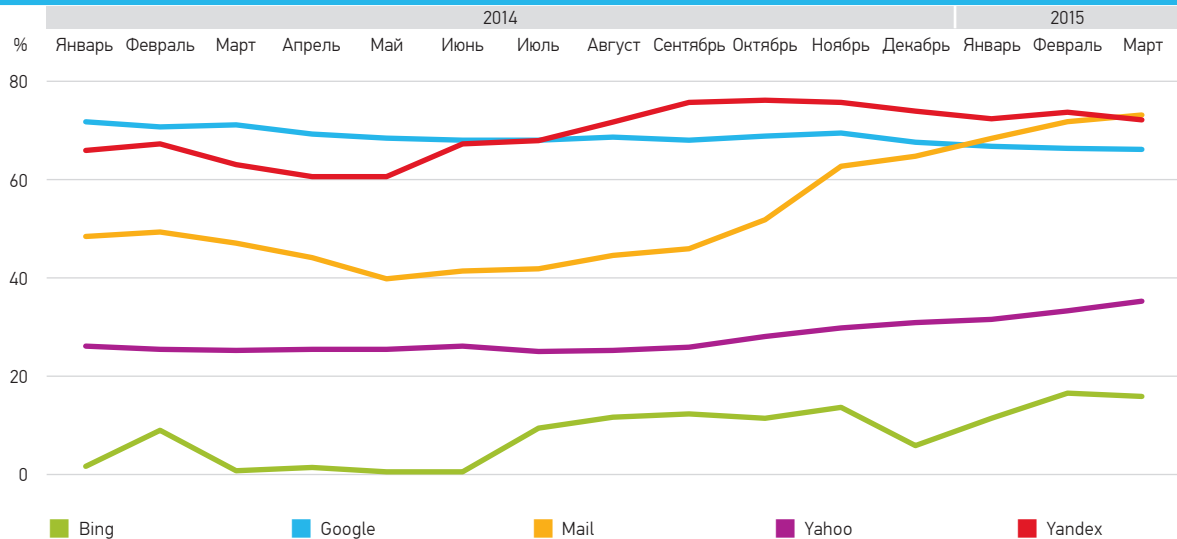
АНАЛИЗАТОР НЕРАЗРЫВНЫХ СЛОВСОЧЕТАНИЙ

Словосочетания в поисковых запросах требуют особой аккуратности при обработке поисковиками. С одной стороны, иногда отдельные слова полезно игнорировать, с другой — необходимо уметь распознавать сочетания слов, которые нельзя разделять.

Запросами этого анализатора являются словосочетания, которые имеют конкретный смысл, теряющийся при исключении любого из слов (например, «набережная часов»). Анализатор оценивает, насколько часто поисковикам удается сохранить словосочетание, а с ним и смысл запроса.

Уже второй год Google остается лидером по сохранению целостности неразрывных словосочетаний, при этом улучшая свои результаты (рис. 29). Остальные поисковики отстают на 10—15 пунктов.

Рисунок 30. Качество поиска имен

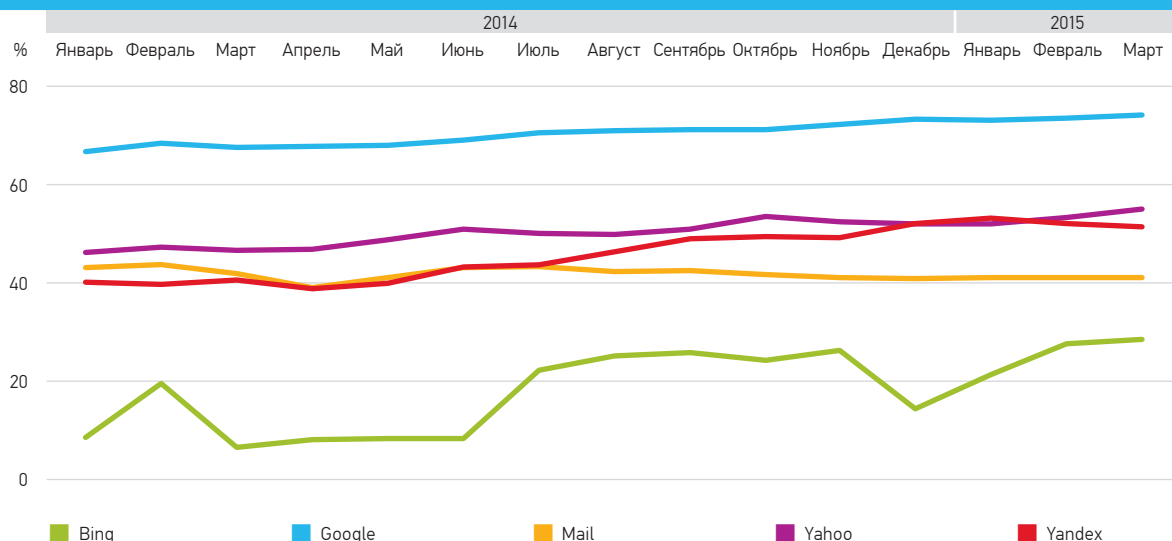


АНАЛИЗАТОР ПОИСКА ИМЕН

Сочетание имени и фамилии — частный случай запроса, который нельзя разбить на части, не потеряв смысл, но мы выделили подобные запросы в отдельный анализатор, т.к. это все же особенный случай. Нужный человек, обнаруженный в результатах поиска, — это плюс для поисковика, а найденные однофамильцы, тезки и пр. — очевидно, минус.

Основным результатом 2014 — начала 2015 года в этом аспекте поиска стало то, что Яндекс и Mail.ru обошли по этому показателю Google (рис. 30). Теперь все три поисковика находят нужных людей примерно на 7 страницах из 10.

Рисунок 31. Качество различения форм слов в поиске



46

АНАЛИЗАТОР ГРАММАТИКИ

Анализатор различения форм слов — один из самых сложных, среди созданных нами. Он был запущен в конце февраля 2013 года, чтобы проиллюстрировать, что чем умнее поисковая машина, тем больше в ней будет и ошибок, вызванных ошибочной трактовкой запроса.

Конечно, в анализаторе проверяется не полная грамматика всего запроса, а лишь соответствие найденной конструкции исходной. Например, по запросу «мехи из кожи» пользователь точно не хочет увидеть сайт про меховые шубы и дубленки, а поисковики, вопреки грамматическому здравому смыслу, все равно добавляют эти сайты в результаты поиска.

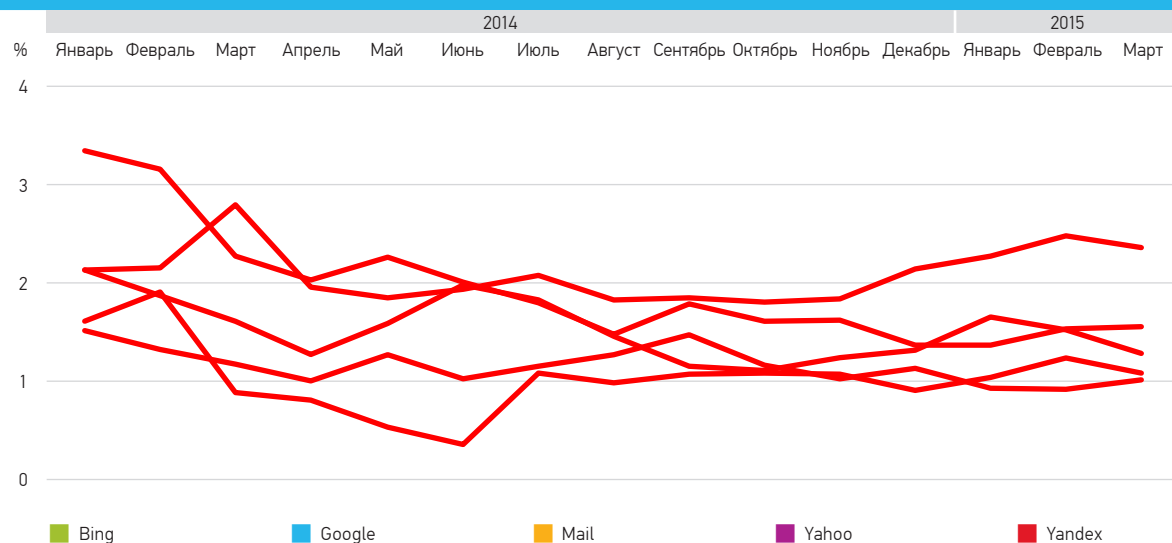
Как и в остальных случаях, когда речь идет об ошибках поисковиков, Google показывает результаты лучше, чем у конкурентов, более 70% правильно учтенных форм слов (рис. 31). Яндекс и Yahoo (около 50%) улучшали результаты в течение года, и только Mail.ru остался на уровне 40%.

10.

РАЗДРАЖАЮЩИЕ ФАКТОРЫ

В ТО ВРЕМЯ КАК СМЫСЛОВЫЕ ОШИБКИ ПОИСКОВИКА МОГУТ КАЗАТЬСЯ ПОЛЬЗОВАТЕЛЯМ ЗАБАВНЫМИ, ЕСТЬ ВЕЩИ, КОТОРЫЕ НЕ ВЫЗЫВАЮТ НИЧЕГО, КРОМЕ РАЗДРАЖЕНИЯ. РЕЧЬ ИДЕТ В ПЕРВУЮ ОЧЕРЕДЬ О ПОКАЗЕ ЗАВЕДОМО БЕССМЫСЛЕННОГО, НЕПРИЯТНОГО ИЛИ ДАЖЕ ОПАСНОГО КОНТЕНТА ИЛИ ССЫЛОК НА НЕГО В РЕЗУЛЬТАТАХ ПОИСКА. ДЛЯ РАБОТЫ НЕСКОЛЬКИХ АНАЛИЗАТОРОВ ЭТОЙ ГРУППЫ ИСПОЛЬЗУЮТСЯ СПЕЦИАЛЬНЫЕ ТЕХНОЛОГИИ, РАЗРАБОТАННЫЕ КОМПАНИЕЙ «АШМАНОВ И ПАРТНЕРЫ», НАПРИМЕР, РАСПОЗНАВАНИЕ И КЛАССИФИКАЦИЯ ПОИСКОВОГО СПАМА.

Рисунок 32. Уровень поискового спама



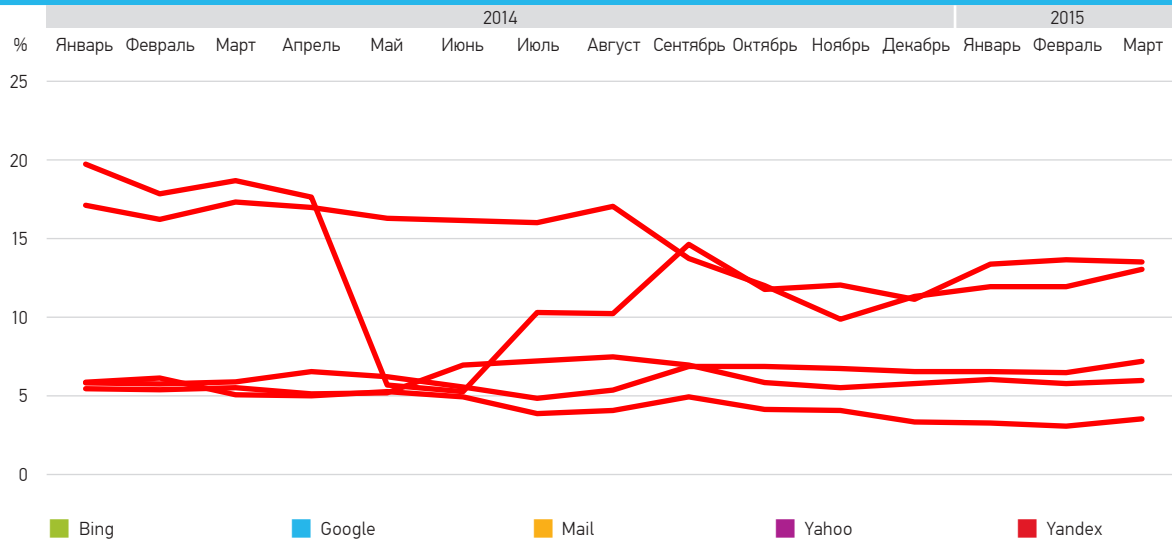
48

АНАЛИЗАТОР СПАМА В ПОИСКОВОЙ ВЫДАЧЕ

Типов поискового спама достаточно много (например, дорвей, спам-каталог, псевдосайт и т. п.). Они различаются используемыми методами, но все их объединяет то, что они созданы в расчете на роботов поисковых машин, а не на людей: пользователю зачастую не удастся понять, что страница бесполезна, пока он не перейдет на сайт, который таким образом получает трафик из результатов поиска.

Неровный характер графиков (рис. 32) говорит о постоянной борьбе, идущей между создателями спама и поисковиками. Тем не менее, один спам-сайт из 100 (у Google) и тем более 2 из 100 у Mail.ru (и это только в первых страницах выдачи) — это довольно много, и оставляет поле для работы.

Рисунок 33. Количество «сайтов для взрослых» по неоднозначным запросам

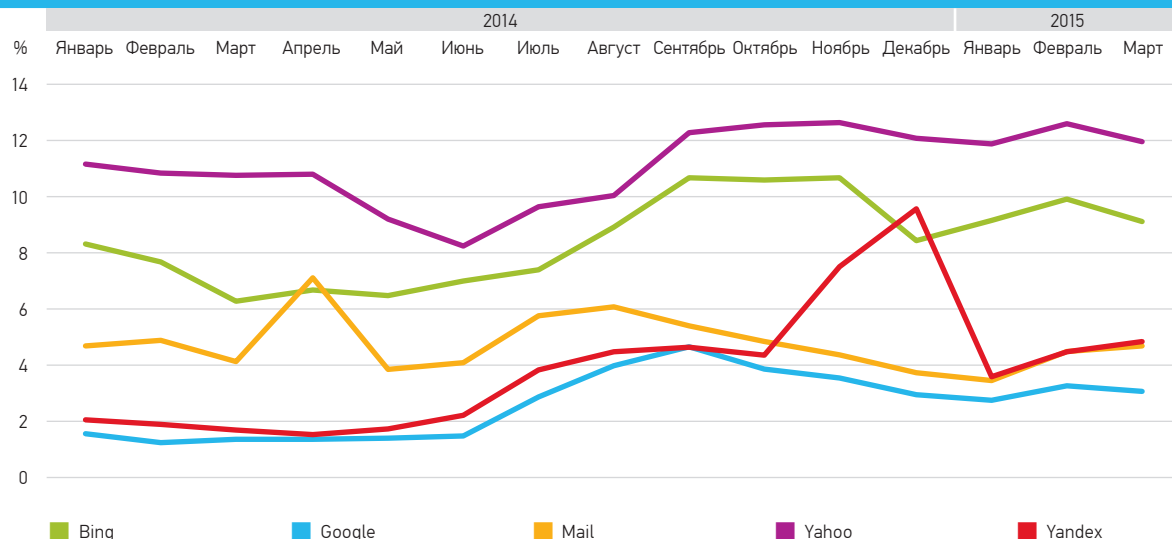


АНАЛИЗАТОР «САЙТОВ ДЛЯ ВЗРОСЛЫХ» В ПОИСКОВОЙ ВЫДАЧЕ

Порносайты или просто страницы, содержащие «контент для взрослых», в результатах поиска по запросам, явно не указывающим на поиск порнографии, могут вызвать у пользователей недоумение и раздражение даже большее, чем просто бессмысленные сайты. Например, запросы «знаменитости», «наручники», «пьяные» и т. п. не могут считаться порно-запросами, и никакого «взрослого» контента по ним находиться не должно.

Год назад мы отмечали, что ситуация с «сайтами для взрослых» в поисковой выдаче далека от идеальной. К началу 2015 года ситуация почти не изменилась: в выдаче Яндекса и Google примерно 5-6 страниц из 100 содержат порнографический контент. Mail.ru удалось за год снизить долю таких страниц до 4% и выйти на первое место по этому показателю (рис. 33).

Рисунок 34. Количество сайтов для взрослых в результатах поиска при включенном семейном фильтре



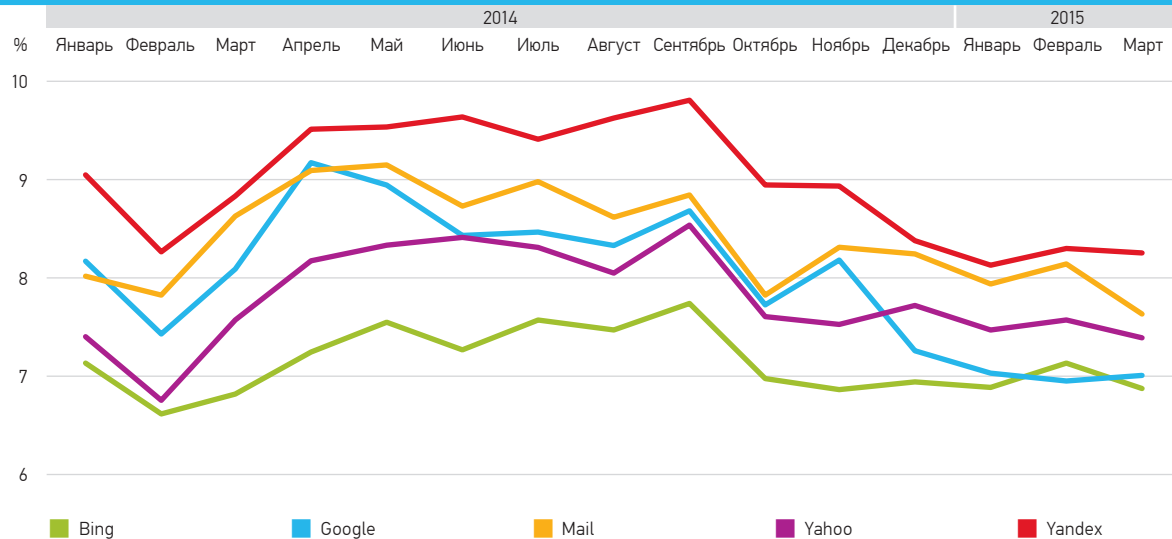
50

АНАЛИЗАТОР КАЧЕСТВА СЕМЕЙНОГО ФИЛЬТРА

Во всех поисковых машинах есть возможность включить фильтрацию выдачи от «плохих сайтов», т.е. в первую очередь, сайтов, не предназначенных для показа детям, — это так называемые семейные фильтры. В идеале результатом работы такого фильтра должно быть полное отсутствие порносайтов в результатах поиска, независимо от запроса. Здесь это особенно важно, ведь семейный фильтр могут использовать родители, чтобы оградить детей от нежелательного содержания страниц.

К сожалению, ситуация и здесь далека от идеала, а динамика, показанная лидерами в 2014 году, была отрицательной (рис. 34). В результате 4—5% страниц, не подходящих для детей, все еще попадают в выдачу даже сквозь семейный фильтр. Стоит отметить, что критерии «плохих» страниц этого анализатора более строгие, чем у простого анализатора порнографии, поэтому и результаты могут быть выше.

Рисунок 35. Уровень рекламной нагрузки на результатах поиска

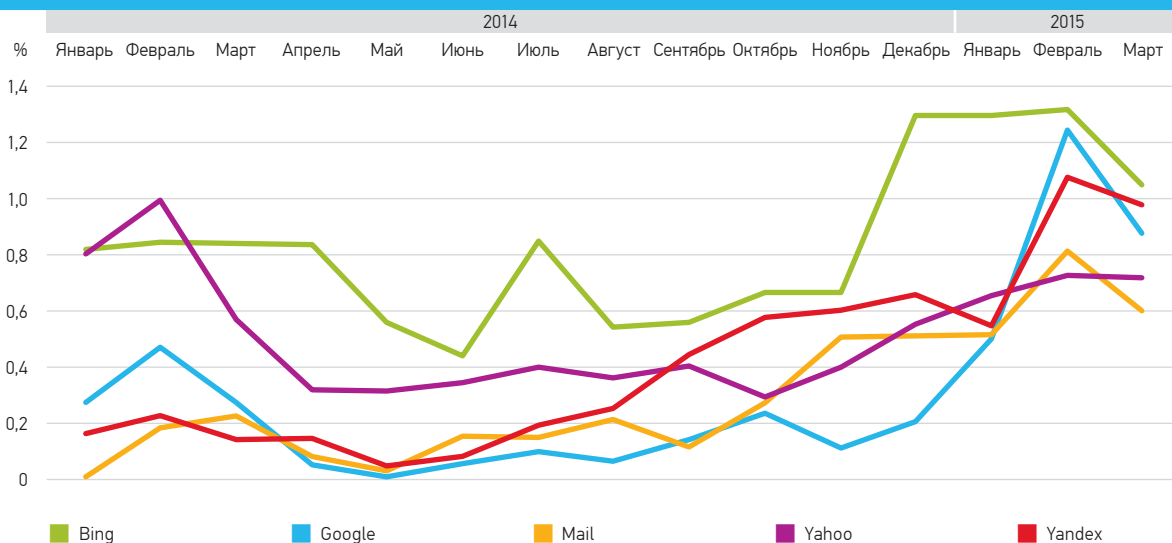


АНАЛИЗАТОР РЕКЛАМНОЙ НАГРУЗКИ

Разработанная нами технология распознавания рекламы на сайтах позволяет автоматически оценить загруженность рекламой всех найденных страниц и выставить каждой из них некоторый балл. Учитывается как навязчивость рекламы, так и ее расположение на странице. Чем сильнее страница перегружена рекламой, тем больше баллов она получит.

К концу 2014 — началу 2015 года поисковики смогли практически сравняться по степени загруженности результатов поиска рекламой (рис. 35). Стоит отметить, что сайты в выдаче Google наименее загружены рекламой (в среднем 7 баллов), а в результатах Яндекса — наоборот (в среднем 8,5 баллов).

Рисунок 36. Количество порнорекламы на страницах из результатов поиска



52

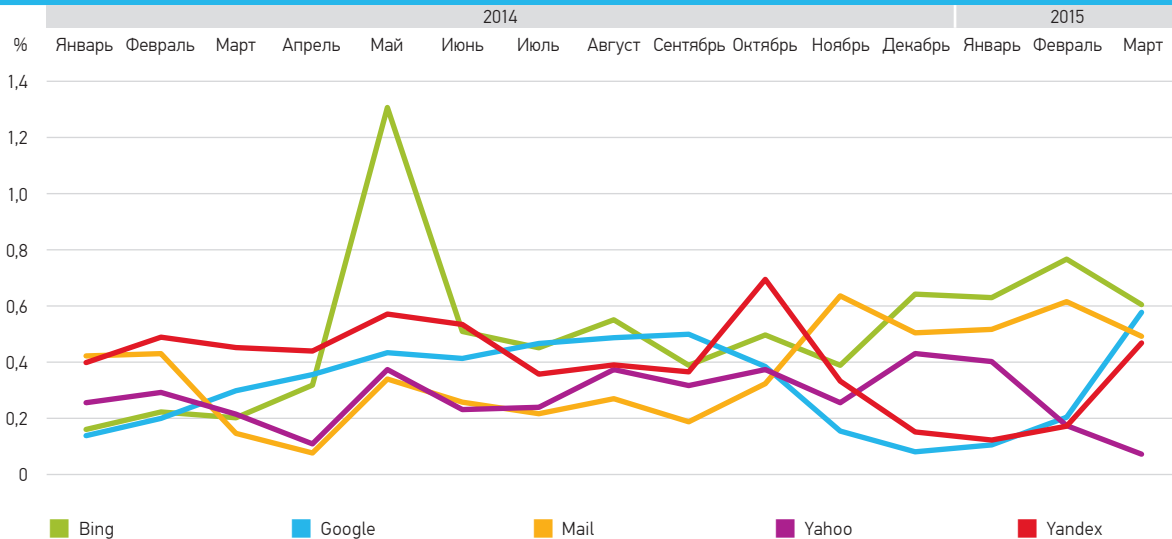
АНАЛИЗАТОР ПОРНОРЕКЛАМЫ

По тем же запросам, которые используются для анализатора рекламной нагрузки, на найденных страницах дополнительно оценивается количество наиболее раздражающей пользователей рекламы — порнографической.

Количество такой рекламы неожиданно резко выросло во всех поисковых системах в конце 2014 — начале 2015 года (рис. 36). Вероятно, это связано с распространением обменных тизерных сетей, которые используют очень откровенные картинки, переходящие в разряд порнорекламы.

Интересно отметить различия между результатами анализаторов общей рекламной нагрузки и порнорекламы. Например, Mail.ru обгоняет Google по степени загруженности выдачи рекламой, но порнорекламы на этих страницах заметно меньше, и поисковик занимает по этому критерию первое место.

Рисунок 37. Доля сайтов с вирусными угрозами в результатах поиска



АНАЛИЗАТОР ВИРУСНЫХ УГРОЗ

Если из результатов поиска пользователь может попасть на сайт, зараженный вирусом, это уже не просто неприятно, но и опасно. Созданный нами анализатор вирусных угроз оценивает количество зараженных сайтов, которые показываются в результатах поиска.

Конечно, все поисковики стараются не показывать на первой странице выдачи сайты, которые могут нанести вред компьютеру пользователя. Доля таких сайтов у всех поисковиков меньше 0,6%, т.е. угрозу заражения содержат примерно 6 страниц выдачи

из 1000, а в конце 2014 года Google удавалось снизить это значение до 1 страницы из 1000 (рис. 37). На графиках можно заметить, насколько нестабильно количество найденных зараженных сайтов: это говорит об очень высокой активности вирусной индустрии, стремящейся постоянно находить новые сайты-жертвы для заражения и о борьбе, которую непрерывно ведут с ней поисковики.

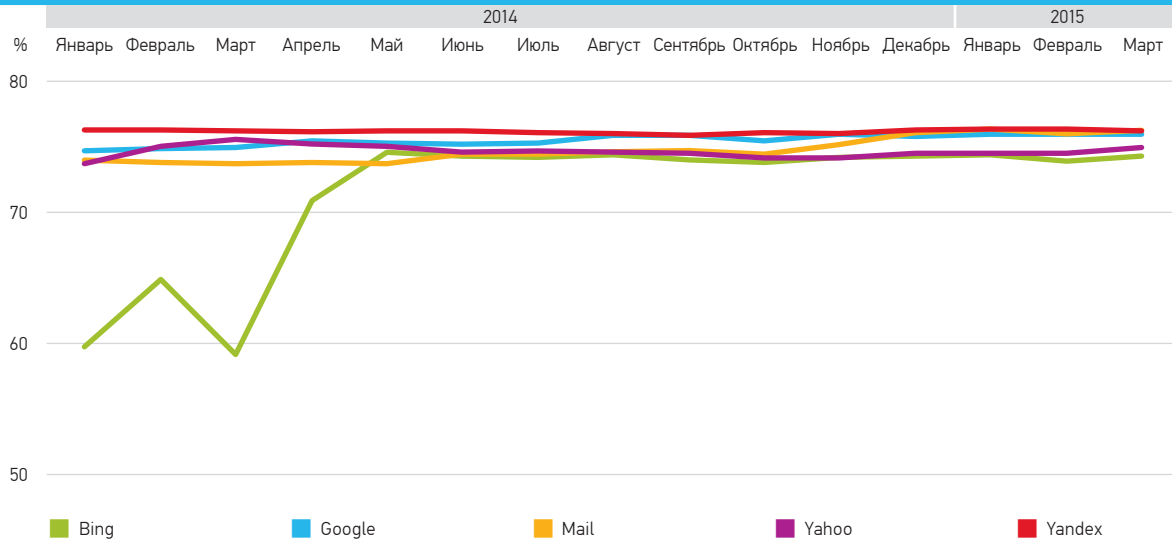
11.

ОБЩИЕ ХАРАКТЕРИСТИКИ ПОИСКА

54

В ЭТОМ РАЗДЕЛЕ МЫ РАССМОТРИМ САМЫЕ РАЗНЫЕ И ПРИ ЭТОМ ОСОБЕННО ВАЖНЫЕ ПОКАЗАТЕЛИ КАЧЕСТВА ПОИСКА, КОТОРЫЕ ОЦЕНИВАЮТ НАШИ АНАЛИЗАТОРЫ. РЕЧЬ ПОЙДЕТ КАК О ТЕХНИЧЕСКИХ ХАРАКТЕРИСТИКАХ, ТАК И ОБ ОЦЕНОЧНЫХ.

Рисунок 38. Ассессорские оценки качества поиска



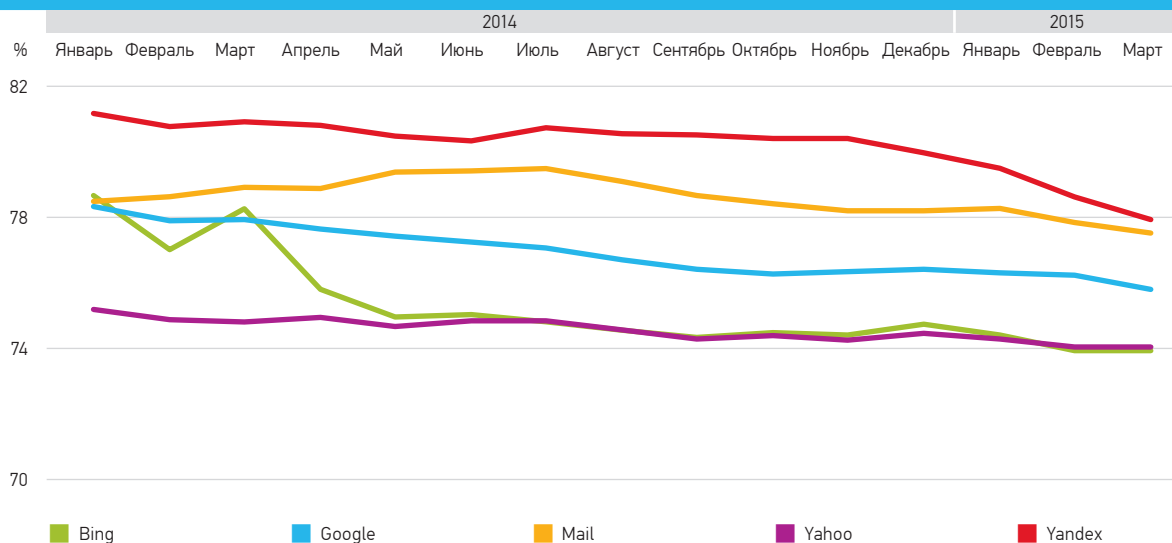
АСЕССОРСКИЙ АНАЛИЗАТОР

Ассессоры — это люди, которые оценивают результаты поиска по специальной шкале, руководствуясь строгой инструкцией (ассессоры работают в штате каждой поисковой машины). В проекте Analyzethis.ru также работают такие специалисты, которые ежедневно выставляют оценки результатам поиска по набору запросов ассессорского анализатора.

Ассессорский анализатор показывает усредненную оценку за релевантность найденных по каждому запросу страниц с учетом качества сайта. К этому могут добавляться штрафы за отсутствие нужных сайтов в выдаче по навигационным запросам.

Наши ассессоры оценивают качество работы трех основных поисковиков примерно одинаково, все оценки находятся в интервале 74—76% (рис. 38). Но если в начале 2014 года оценки Яндекса были пусть и немного, но заметно выше остальных, то к началу 2015 года качество всех трех основных поисковиков практически сравнялось, Mail.ru и Google догнали Яндекс по уровню релевантности найденных страниц.

Рисунок 39. Качество снипетов



56

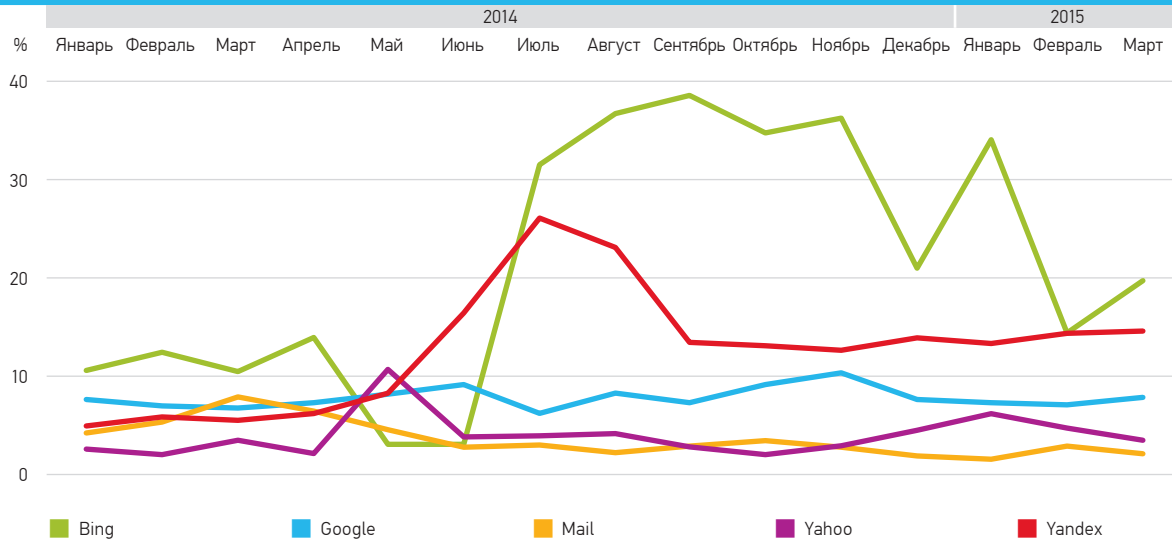
КАЧЕСТВО СНИПЕТОВ

Снипеты — это то, что видит пользователь в результатах поиска и на основании чего он решает, на какую страницу ему перейти. То есть этот набор из заголовка страницы и кусочка текста, подобранный поисковиком, играет очень заметную роль.

Качество снипетов оценивают ассесоры одновременно с оценкой релевантности поиска в ассесорском анализаторе. На оценку снипета влияют как его соответствие содержанию страницы, так и его качество и состав.

Как и в 2013, в прошедшем году качество снипетов постепенно снижалось у всех поисковиков: у Яндекса его оценка упала с 81% до 78%, у Google — с 78% до 76% (рис. 39). Отметим, что по качеству снипетов поиск Mail.ru занимает уверенное второе место, обгоняя Google.

Рисунок 40. Средний уровень апдейтов за месяц



АНАЛИЗАТОР АПДЕЙТОВ

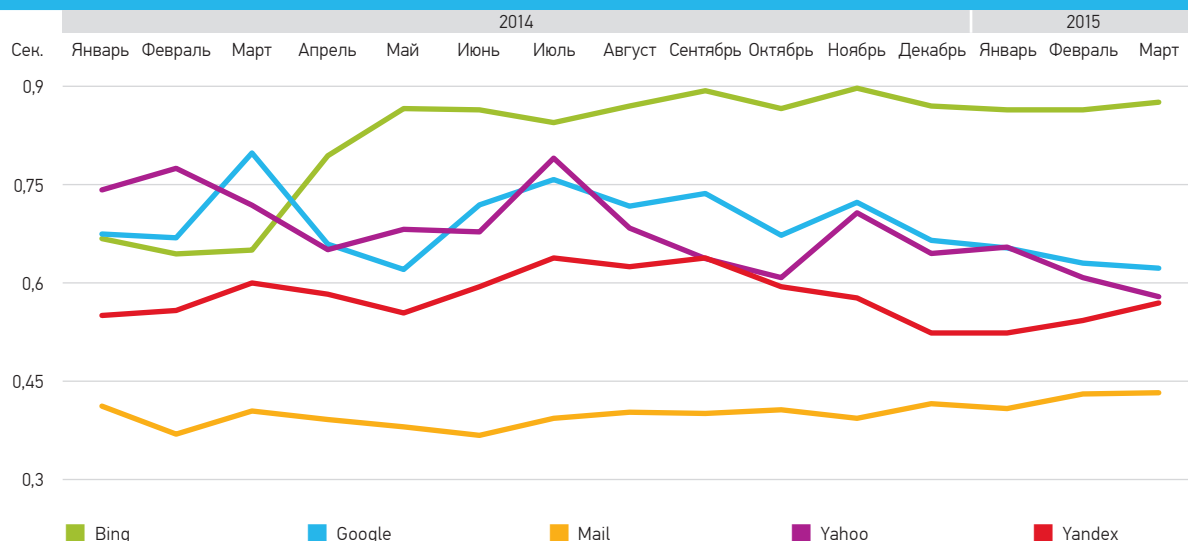
Апдейт — это обновление индекса поисковика либо изменение в алгоритмах ранжирования, за которыми следует и изменение результатов поиска по отдельным запросам.

Любопытно, что у каждой поисковой машины есть свой характер проведения обновлений, что видно на среднем ежемесячном графике апдейтов (рис. 40).

Например, по сравнению с 2013 годом характер обновлений в поиске Mail.ru стал более плавным и спокойным, в то время как Яндекс, наоборот, вносил резкие изменения в июне-августе 2014 года, когда свои позиции в результатах меняли до четверти всех страниц.

Отдельный интерес представляет и график ежедневных апдейтов, доступный на сайте www.analyzethis.ru.

Рисунок 41. Скорость регионального поиска



58

АНАЛИЗАТОР СКОРОСТИ РЕГИОНАЛЬНОГО ПОИСКА

Для анализатора регионального поиска мы задаем поисковые запросы из разных российских городов: Владивостока, Санкт-Петербурга, Самары и других. При этом мы измеряем время, за которое местные серверы в этих городах получают результаты поиска, чтобы оценить, сколько времени придется пользователям ждать ответа от поисковиков в различных регионах.

Результат анализатора скорости регионального поиска — усредненная по всем городам и запросам скорость загрузки страницы поисковой выдачи. При этом на странице анализатора на сайте www.analyzethis.ru доступны всевозможные детали, касающиеся распределения скорости загрузки страниц

по городам и запросам (например, во Владивостоке загрузка результатов поиска занимает почти секунду, а в Google — даже больше).

Поисковики успевают выдать результаты в среднем меньше, чем за 0,7 секунды (рис. 41). Самым быстрым же оказался поиск Mail.ru, вероятно, обладающий не такой сложной серверной структурой, как у Google и Яндекса.

ЧАСТЬ 2

1.

ОСНОВНЫЕ СОБЫТИЯ НА РЫНКЕ ПОИСКА В РУНЕТЕ

В 2014 — I КВАРТАЛЕ 2015 ГОДА

ОБЩИЕ ТЕНДЕНЦИИ РАЗВИТИЯ РУНЕТА И ВАЖНЕЙШИЕ СВЯЗАННЫЕ С НИМ СОБЫТИЯ ТРАДИЦИОННО СКАЗАЛИСЬ И НА ИГРОКАХ РЫНКА ПОИСКА.

1.1. ЭВОЛЮЦИЯ ЧИСЛЕННОСТИ, СОСТАВА И ПОВЕДЕНИЯ РОССИЙСКОЙ ИНТЕРНЕТ- АУДИТОРИИ

Количество людей, которым доступен интернет и которые пользуются связанными с ним продуктами и услугами в России продолжает увеличиваться. Аудитория прирастает жителями регионов, относительно небольших населенных пунктов, представителями старших возрастных групп. Именно эти категории будут в ближайшее время обеспечивать рост Рунета, поскольку молодые

и среднего возраста люди из крупных городов давно стали активными пользователями.

Для интернет-компаний это может иметь несколько последствий. Во-первых, появляются новые социально-демографические группы потенциальной аудитории нишевых сервисов (продукты для детей, для пенсионеров и так далее). Во-вторых, компании с массовыми, ориентированными на широкую аудиторию сервисами (например, поисковые системы) вынуждены заботиться об адаптации своих разработок и интерфейсов под нужды самых разных людей: пользователей медленного или дорогого интернет-соединения, несовершеннолетних, которых нужно защищать от опасного контента, молодых людей, которые еще не знакомы со способами интернет-мошенничества и другими типичными сетевыми угрозами.

1.2. РОСТ ПОПУЛЯРНОСТИ МОБИЛЬНЫХ УСТРОЙСТВ

Исследовательская компания TNS обнаружила, что в 2014 г. количество пользователей, выходящих в интернет с настольных компьютеров (десктопов), впервые за всю историю измерений не выросло. Более того, в 2015 году компания заметила снижение количества пользователей с десктопами. Все больше людей предпочитают выходить в Сеть с мобильных устройств — планшетов и смартфонов. Можно выделить несколько основных последствий для интернет-компаний. Меняется поведение пользователей: при заходе с мобильного устройства они проводят в интернете гораздо меньше времени за одну сессию, активно используют в первую очередь навигационные и коммуникационные сервисы, предпочитают более простые интерфейсы, реже делают покупки. На мобильных устройствах обостряется война за внимание пользователя. Согласно исследованиям, относительно небольшая доля владельцев смартфонов и планшетов устанавливают даже бесплатные приложения, не говоря о покупке платных; большинство установленных приложений запускаются несколько раз и забрасываются. Добиться того, чтобы приложение заинтересовало пользователя в магазине, тоже оказывается непросто или дорого для многих компаний и разработчиков.

С ростом количества устройств, которыми владеет население, распространяется мультиэкранное потребление информации. Разработчикам и маркетологам теперь нужно учитывать ситуации, когда пользователь начинает решать свою задачу на одном устройстве, а заканчивает на другом, либо пользуется несколькими параллельно. Ярким примером такой ситуации стало вручение премии «Оскар» в 2013–2015 году: люди смотрели церемонию по телевизору и одновременно искали в интернете платья знаменитостей (пытаясь купить себе такие же), подробности биографии актеров и другую связанную с шоу информацию, а также рассказывали о своих впечатлениях по ходу просмотра в социальных сетях.

Наконец одно из важнейших для поискового рынка последствий «мобилизации» аудитории — это давление на рынок системы дистрибуции продуктов, привязанной к мобильным операционным системам. Огромная доля приобретенных российскими пользователями устройств работает на операционной системе Android, принадлежащей Google. Более подробную информацию о том, как это влияет на компанию Яндекс, вы найдете в соответствующем разделе.

1.3. УСИЛЕНИЕ ГОСУДАРСТВЕННОГО РЕГУЛИРОВАНИЯ

В 2014 году было принято несколько важных законов и предложено много инициатив, меняющих жизнь интернет-компаний. Закон, обязывающий хранить персональные данные российских пользователей на территории России, вступает в силу только с 1 сентября 2015 года, но компании уже сейчас ищут способы ему соответствовать, не нарушая доступности собственных сервисов для жителей РФ.

Усиление контроля за поведением популярных блогеров в интернете, возможно, повлияло на решение Яндекса закрыть собственную блогговую платформу. Случились и другие подвижки: внесенные в Гражданский Кодекс поправки упростили сделки по венчурному инвестированию.

1.4. НЕБЛАГОПРИЯТНАЯ ЭКОНОМИЧЕСКАЯ СИТУАЦИЯ

Экономический кризис и околполитические события не могли не сказаться и на бизнесе интернет-компаний. Рост доллара к рублю удорожил для Яндекса аренду его московского офиса, закупку серверов и строительство дата-центра в Финляндии. Согласно финансовым итогам

2014 года у Яндекса упала выручка от медийной рекламы, хотя контекст продолжал расти. В первом квартале 2015 года Яндекс ожидает снижение рентабельности. Компания произвела оптимизацию штата — часть сотрудников была сокращена, бюджет и стратегия найма новых кадров пересмотрены. От кризиса пострадали и бизнесы, покупающие контекстную рекламу в поисковых системах. Случился спад активности у рекламодателей в таких сегментах как банки, авто, туризм.

1.5. КОНКУРЕНЦИЯ И СОТРУДНИЧЕСТВО С СОЦИАЛЬНЫМИ СЕТЯМИ

Поисковики и социальные сети являются самыми массовыми сервисами в Интернете с огромным количеством пользователей и информации о них. Исторически они не конкурировали за внимание людей, так как решали разные их задачи. Однако социальные сети постепенно превращаются из инструмента коммуникации в полноценную среду существования с внутренними источниками информации, развлечений, услуг, медийного контента. Навигацию по этому контенту они стремятся обеспечивать своими силами и с учетом собственных интересов.

Поисковику, в свою очередь, нужен контент из социальных сетей и информация о пользователях.

Между поисковыми системами и соцсетями складываются непростые отношения сотрудничества и конкуренции. В 2014–начале 2015 года самыми важными событиями стали следующие:

- ▶ Яндекс договорился с Facebook об индексации содержимого этой соцсети и отображению его в результатах поиска.

- ▶ Facebook фактически отключил появившийся в 2013 году поисковый алгоритм Graph Search, который отвечал на вопросы на естественном языке, но не искал по тексту постов. В 2014 году заработал новый поисковый алгоритм, который умеет искать по тексту собственных постов пользователя и записей его друзей. Партнерство с поисковику Bing от Microsoft было расторгнуто.
- ▶ ВКонтакте начал экспериментировать с показом на своих страницах рекламы Яндекс.Директа.
- ▶ Google снова заключил с Twitter договор о доступе к Firehose — потоку всех твитов в режиме реального времени. Компании уже сотрудничали раньше, затем Twitter прервал отношения с Google, в 2015 сотрудничество возобновилось. Теперь в поисковой выдаче Google будут появляться самые свежие твиты, что, например, будет полезно пользователю, ищущему информацию об актуальных событиях.

2.

ЯНДЕКС В 2014 — I КВАРТАЛЕ 2015 ГОДА

62

ЭТОТ ГОД ОКАЗАЛСЯ КРАЙНЕ СЛОЖНЫМ для ЯНДЕКСА, ХОТЯ СЛЕДУЮЩИЙ ПЕРИОД ОБЕЩАЕТ БЫТЬ ЕЩЕ СЛОЖНЕЕ. СТАЛИ ПОНЯТНЫ ГЛОБАЛЬНЫЕ ПРОБЛЕМЫ КОМПАНИИ И СПОСОБЫ ИХ РЕШЕНИЯ, НА КОТОРЫЕ ЯНДЕКС СДЕЛАЕТ СТАВКУ.

Экономический кризис заставляет Яндекс повышать собственную финансовую эффективность. Доминирование Google на рынке мобильных платформ требует ужесточение борьбы за внимание пользователей устройств.

Поисковая компания продолжает реализацию своей новой стратегии, анонсированной несколько лет назад: она теперь не ищет ответы на вопросы, а помогает пользователям решать их задачи. Эта стратегия влияет и на запуски новых сервисов, и на развитие уже существующих.

2.1. БОРЬБА С GOOGLE ЗА МОБИЛЬНУЮ АУДИТОРИЮ

В феврале 2014 года Яндекс представил собственную прошивку для устройств на Android под названием Яндекс.Кит. В марте в продажу поступили первые смартфоны на Яндекс.Ките — Explay Flame и Honor 3 Yandex Huawei. Однако осенью того же года Explay, занимающий 5,2% рынка смартфонов в России, отказался от сотрудничества с Яндексом и начал работать по соглашению с Google.

Зимой 2015 года Яндекс подал в Федеральную Антимонопольную Службу России заявление с просьбой проверить деятельность Google на соответствие антимонопольному законодательству. Google использует популярность своей операционной системы Android для продвижения других своих сервисов и ограничения

конкуренции с ними чужих продуктов, — объясняют свои претензии представители Яндекса.

В конце 2014 года статистика Liveinternet впервые за 4 года зафиксировала падение доли поиска Яндекса в Рунете ниже 60%. С начала 2014 года количество переходов с Google на сайты Рунета плавно росло, а с Яндекса — падало. Это объясняется ростом использования мобильных устройств для выхода в интернет. Большинство из них работают на OS Android и содержат предустановленные сервисы Google. Пользователи предпочитают запускать уже имеющиеся на устройстве приложения, а не искать и скачивать новые. Таким образом, возможность договориться с производителем о предустановке своих сервисов на первый экран его устройств является критически важной для работы с мобильными пользователями. Google лишает Яндекс такой возможности, согласно заявлению последнего. Если Яндексу не удастся повысить доступность своих продуктов для владельцев мобильных устройств, падение его доли продолжится.

О важности влияния предустановок на популярность сервисов гендиректор Яндекса Александр Шульгин рассказал «Коммерсанту» следующее: «...доля поиска «Яндекса» на Android в целом в четвертом квартале 2014 года — 44%, при этом доля «Яндекса» на Nokia X, где мы установлены поиском по умолчанию, — 76%. Когда на iOS нельзя было поставить поиском по умолчанию «Яндекс», наша доля на этой платформе составляла около 35%. После включения «Яндекса» в список возможных поисковиков по умолчанию наша доля выросла до 49% по итогам четвертого квартала 2014 года».

2.2. УЖЕСТОЧЕНИЕ БОРЬБЫ С МАНИПУЛЯЦИЯМИ ПОИСКОМ

Почти вся история Яндекса является историей противостояния с серыми оптимизаторами, пытающимися

влиять разными методами на его поисковые результаты в собственных коммерческих интересах. Однако в последние полтора года методы противостояния заметно ужесточились.

В апреле 2015 года был анонсирован новый поисковый алгоритм «Минусинск», который включился в мае. Алгоритм призван понизить в ранжировании сайты, замеченные в массовой закупке SEO-ссылок для влияния на свои позиции в поиске. Вебмастерам был дан месяц на то, чтобы снять «неестественные» ссылки.

Ранее Яндекс не наказывал сайты за покупку ссылок, утверждая, что если сайт качественный и решает проблему пользователя, он должен присутствовать в результатах поиска. Кроме того, в компании утверждают, что поисковые алгоритмы способны распознавать SEO-ссылки и самостоятельно нейтрализовывать их влияние на ранжирование страниц. Борьба с их закупкой, Яндекс собирается научить компании правильно тратить деньги — не на бесполезные ссылки, а на дизайн, юзабилити, контент и другие инструменты повышения качества ресурсов.

В марте 2014 года Яндекс заявил о том, что поисковая выдача для пользователей из Москвы по коммерческим запросам теперь формируется вообще без учета ссылок между сайтами. Для оценки последствий этого шага на сайте analyzethis.ru появился новый анализатор Коммерческих апдейтов. Он показал, что по соответствующим нововведению запросам выдача изменилась заметно, но не глобально. Это говорит о том, что новая технология ранжирования запускается очень постепенно и была тщательно подготовлена.

Однако многие оптимизаторы считают, что реальные меры по борьбе с неестественными ссылками, предлагаемые Яндексом, гораздо мягче и осторожнее, чем декларируемые. Ссылочные связи страниц

в интернете слишком долго являлись основным инструментом поискового ранжирования, и отказ от них может привести к последствиям, которые отпугнут пользователей.

В 2014 году Яндекс наказывал сайты еще и за накрутку поведенческих факторов — то есть, за применение технологий, симулирующих использование сайта живыми людьми. Поведение пользователей на сайте Яндекс отслеживает с целью понимания его качества, удобства, способности решать задачи пользователя, релевантности запросу и других важных параметров.

2.3. РАЗВИТИЕ АЛЬТЕРНАТИВНЫХ МОДЕЛЕЙ ПОЛУЧЕНИЯ ПРИБЫЛИ, РАБОТА С БОЛЬШИМИ ДАННЫМИ

64

Основным источником доходов Яндекса исторически являлась контекстная реклама. Этот рынок продолжает расти несмотря на экономический кризис, однако Яндекс ищет другие способы зарабатывать деньги.

Помимо Яндекс.Такси, модель комиссии от партнеров по предоставлению пользователям платных услуг задействовал Яндекс.Мастер (master.yandex.ru). За год у компании появились три платных API — карты, перевод, голосовые технологии. На «Кинопоиске» Яндекс собирается показывать фильмы, зарабатывая на рекламе или подписке.

Но главное — это запуск в декабре 2014 года проекта Yandex Data Factory. Для компаний, которые располагают большими массивами данных, Яндекс будет эти данные обрабатывать и искать таким образом решение поставленных задач. Для этого компания задействует свои ключевые технологии — машинное обучение, распознавание образов и речи, глубокие нейронные сети, обработку естественного языка.

К весне 2015 года Data Factory уже реализовала несколько проектов. Для «Росавтодора» была разработана система прогнозирования пробок и ДТП. Для разработчика игр Wargaming Яндекс предсказывает периоды оттока аудитории — сколько людей и когда перестанут играть. Аналогичную услугу — прогнозирование оттока абонентов — Яндекс оказывает южноафриканскому оператору мобильной связи. В рамках сотрудничества с ЦЕРН Yandex Data Factory анализирует данные, собранные аппаратурой центра ядерных исследований.

2.4. ЗАМОРОЗКА ИЛИ ПЕРЕЗАПУСК «ОСТРОВОВ»

В 2013 году Яндекс анонсировал концепцию «Островов» — интерактивной поисковой выдачи, в которой можно взаимодействовать с сайтами, не переходя на них. Подразумевалось, что пользователь сможет, например, записаться в парикмахерскую, купить билет, получить нужную информацию прямо в поисковой выдаче, если вебмастер реализует для своего сайта соответствующую функцию.

В 2014 году Яндекс два раза анонсировал запуск новой версии поиска с «Островами», однако оба раза поисковик испытывал технические проблемы с доступностью сайта и качеством выдачи.

В апреле 2015 вебмастера обнаружили, что инструментарий работы с «Островами» недоступен. Яндекс прокомментировал этот факт заявлением, что от концепции интерактивных ответов в поиске он не отказывается, и что через некоторое время для вебмастеров появятся новые инструменты.

2.5. АКЦЕНТ НА РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ ПЕРСОНАЛИЗАЦИИ

Есть основания предполагать, что в ближайшее время Яндекс вплотную займется внедрением рекомендательных технологий в разные сервисы. В 2014-начале 2015 года интерес компании к системам рекомендации становился все более заметным. В сентябре 2014 года система персональных рекомендаций была внедрена в приложении Яндекс.Музыка; весной 2015 года основатель рекомендательной системы материалов для СМИ рассказал, что Яндекс разрабатывает аналогичный сервис. Еще в 2013 году была анонсирована платформа Яндекс.Атом, которая должна позволить адаптировать любой сайт под потребности и поведение конкретного пользователя. К весне 2015 года «Атом» находится в режиме закрытого тестирования.

2.6. ЗАКРЫТИЕ СЕРВИСОВ И ПОКУПКА КОМПАНИЙ

За последний год Яндекс стал гораздо более решительно избавляться от недостаточно популярных, непрофильных, медленно развивающихся или по иным причинам найденных неудачными сервисов. В 2015 году были закрыты «Услуги», помогавшие подобрать кредит, ипотеку или вклад; блогговая платформа Я.ру, сервис «Закладки». На Яндекс.Видео пользователи больше не могут загружать свои ролики для публичного просмотра, они могут только хранить личные видеозаписи на Яндекс.Диске. Закрылся рейтинг блогов — как утверждают в компании, в связи с потерей блогами популярности в пользу соцсетей. С главной страницы yandex.ru исчез рейтинг популярных в блогах тем.

При этом в 2015 году Яндекс показал рекордную для себя активность в покупке сторонних компаний. Если в 2013 был приобретен только «Кинопоиск», то к октябрю 2014 — уже семь стартапов.

2.7. ЯНДЕКС В МАССОВОЙ КУЛЬТУРЕ

В 2014 году вышли две книги про Яндекс и фильм «Стартап», в котором название компании не упоминалось, но особенности ее истории легли в основу сюжета. «Яндекс.Книга» была создана под патронажем самой компании, а от фильма «Стартап» Яндекс дистанцировался. Фильм был призван повысить популярность технологического предпринимательства среди россиян и создать положительный образ стартапера.

3.

GOOGLE В 2014 — I КВАРТАЛЕ 2015 ГОДА

66

3.1. РАНЖИРОВАНИЕ САЙТОВ НА ОСНОВЕ ДОСТОВЕРНОСТИ

Весной 2015 года Google опубликовал описание нового способа ранжирования сайтов — «Доверие на основе достоверности», Knowledge-Based Trust. Его авторы предлагают оценивать веб-страницы на основании фактической точности и безошибочности опубликованной на них информации. Они согласны с Яндексом в том, что ранжирование страниц на основании ссылок на них с других страниц неоптимально и уязвимо для манипуляций.

Этот алгоритм применим не к любым документам, а только к тем, что содержат фактическую информацию, достоверность которой может быть оценена. Оценку Google производит путем сопоставления сообщающихся на странице фактов с имеющимися в его собственной базе знаний

3.2. КУРС НА АДАПТАЦИЮ ВСЕХ САЙТОВ ПОД МОБИЛЬНЫЕ УСТРОЙСТВА

Рост интенсивности использования интернета на мобильных устройствах заставил и Google принять определенные меры. Компания объявила, что с апреля 2015 при определении поискового рейтинга сайта будет учитывать, удобно ли просматривать его страницы на смартфонах и планшетах. Для вебмастеров появился инструмент проверки совместимости сайта с мобильными гаджетами.

3.3. ОТКАЗ ОТ РАЗВИТИЯ СОЦИАЛЬНОЙ СЕТИ «КЛАССИЧЕСКОГО» ТИПА И ИНТЕГРАЦИИ ЕЕ С ПОИСКОМ

Весь 2014 год разные эксперты наблюдали и анализировали признаки угасания социальной сети Google+, уменьшения количества выделяемых на нее ресурсов и торможения развития сервиса. Google понижает плотность интеграции Google+ с другими продуктами и сокращает влияние связанных с соцсетью факторов на поисковую выдачу.

В апреле 2014 года стало известно, что возглавлявший работу над соцсетью Вик Гундотра покидает Google. В марте 2015 вице-президент компании сообщил, что Google+ будет «разобрана» на несколько отдельных мобильных приложений — мессенджер, новостную ленту, фоторедактор.

В УКАЗАННЫЙ ПЕРИОД РАЗВИТИЕ ПОИСКА СТАЛО ДЛЯ MAIL.RU GROUP ОДНИМ ИЗ МАЛО-ПРИОРИТЕТНЫХ НАПРАВЛЕНИЙ. КОМПАНИЯ СКОНЦЕНТРИРОВАЛА РЕСУРСЫ НА РАЗВИТИИ МОБИЛЬНЫХ ПРОДУКТОВ ДЛЯ ЗАПАДНОЙ АУДИТОРИИ ПОД БРЕНДОМ MY.COM. ДРУГИЕ ПРИОРИТЕТНЫЕ ОБЛАСТИ — СОЦИАЛЬНЫЕ СЕТИ, ПОЧТА, ИГРЫ И ОБЛАЧНЫЕ СЕРВИСЫ. КРОМЕ ТОГО, КОМПАНИЯ ПОТРАТИЛА ЗНАЧИТЕЛЬНЫЕ СРЕДСТВА НА ПРИОБРЕТЕНИЕ ДОЛИ В СОЦСЕТИ ВКОНТАКТЕ.

За весь 2014 год в блоге Поиска были опубликованы всего четыре новости (например, только в ноябре 2013 было пять сообщений о новых возможностях поиска). Впрочем, степень корреляции количества постов с количеством релизов неизвестна.

Важное обновление поиска в 2014 году: разработчики поиска вручную отобрали несколько сотен сайтов и настроили для них показ в выдаче функциональных кнопок. Среди них, например, сайт радио Energy с кнопкой «слушать», банк «Русский стандарт» с кнопкой входа в онлайн-банк, «Женское такси» с кнопкой «Заказать» и другие.

4.

ПОИСК@MAIL.RU В 2014 — I КВАРТАЛЕ 2015 ГОДА

По нажатию на кнопку пользователь попадает на страницу сайта с соответствующей функцией, формой или информацией, поэтому нововведение Mail.ru — это не аналог Яндекс.Островов. Однако в его основе лежит похожая идея: выдача поисковика не должна быть списком ссылок, она должна быть местом и инструментом решения конкретных задач пользователя.

Доля поиска Mail.ru весь 2014 — 2015 годы постепенно падала по статистике Liveinternet. В январе она составляла 9,2% переходов на сайты Рунета, в апреле 2015 — 6,8%. Общее качество поиска по оценке анализаторов Analyzethis.ru остается примерно равным качеству поиска Яндекса.

5.

«СПУТНИК» В 2014 — I КВАРТАЛЕ 2015 ГОДА

68

В 2014 ГОДУ В РУНЕТЕ ПОЯВИЛАСЬ НОВАЯ ПОИСКОВАЯ СИСТЕМА — ЗАПУЩЕННЫЙ «РОСТЕЛЕКОМОМ» «СПУТНИК». ЕГО ГЛАВНАЯ ЗАДАЧА — «СОЗДАНИЕ ТОЧКИ ДОСТУПА К РАЗВИВАЮЩЕЙСЯ ЦИФРОВОЙ ИНФРАСТРУКТУРЕ ОБЩЕСТВА И ГОСУДАРСТВА», КАК ОПРЕДЕЛИЛ ЕЕ ВИЦЕ-ПРЕЗИДЕНТ «РОСТЕЛЕКОМА» АЛЕКСЕЙ БАСОВ.

Новый поисковик тесно интегрирован с госуслугами и сервисами социального обслуживания. Sputnik.ru помогает найти ближайшие поликлиники, аптеки и другие учреждения, устроиться на работу, записать ребенка в детсад и так далее. Раздел «Удобная Россия» наполнен информацией по взаимодействию с госучреждениями, оформлению документов.

Приоритетом является отсутствие неточной информации и угроз безопасности. По словам Алексея Басова, за четыре года «Спутник» должен войти в топ-10 сайтов Рунета по аудитории.

За первые десять дней после запуска Sputnik.ru сайт посетило 2 млн человек, однако в 2014 году его поисковая доля стабилизировалась на значениях менее 0,01%, в месяц с него совершается приблизительно 2300 переходов на сайты Рунета.

В октябре 2014 года у портала появилась мобильная версия. В феврале 2015 года вышло мобильное приложение — браузер «Спутник». Еще раньше появились приложения «Мой дом» и «Лекарства». В апреле 2015 на кириллическом домене Sputnik.Дети заработал детский сайт с играми, мультфильмами, образовательными материалами и безопасным поиском.

ЧАСТЬ 3

1.

НОВЫЙ ПРОЕКТ: SEO-АНАЛИЗАТОР

ЛЕТОМ 2014 ГОДА КОМПАНИЯ «АШМАНОВ И ПАРТНЕРЫ» ЗАПУСТИЛА НОВЫЙ ПРОЕКТ — ЛАБОРАТОРИЮ ИЗУЧЕНИЯ ФАКТОРОВ РАНЖИРОВАНИЯ. СПЕЦИАЛИСТЫ ЛАБОРАТОРИИ ИЗУЧАЮТ ВЫДАЧУ ВЕДУЩИХ ПОИСКОВИКОВ ПО КОММЕРЧЕСКИМ ЗАПРОСАМ — В ОСНОВНОМ, ВЫСОКОКОНКУРЕНТНЫМ — И ИЩУТ ЗАКОНОМЕРНОСТИ. ОСНОВНАЯ ЗАДАЧА ЛАБОРАТОРИИ — БОЛЕЕ ТОЧНОЕ ОПРЕДЕЛЕНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА РАНЖИРОВАНИЕ РЕЗУЛЬТАТОВ ПОИСКОВИКАМИ, И СТЕПЕНИ ИХ ВЛИЯНИЯ. ПОСКОЛЬКУ ЗАПРОСЫ ИСКЛЮЧИТЕЛЬНО КОММЕРЧЕСКИЕ, НАРЯДУ С ТЕКСТОВЫМИ И ССЫЛОЧНЫМИ, РАССМАТРИВАЮТСЯ ТАКЖЕ ТАК НАЗЫВАЕМЫЕ «КОММЕРЧЕСКИЕ» ФАКТОРЫ.

ЧТО ТАКОЕ КОММЕРЧЕСКИЕ ФАКТОРЫ РАНЖИРОВАНИЯ

Коммерческие факторы — условное название свойств страницы или сайта, так или иначе важных для принятия посетителем сайта решения о том, сделать ли здесь заказ или звонок. К ним относятся удобство сайта, цены (и их наличие), ассортимент, возможности оплаты и доставки, контакты, наличие гарантий и постпродажного обслуживания, отзывы, качественные фотографии и т. д. Известно, что с конца 2011 года «Яндекс» учитывает эти и подобные характеристики при ранжировании сайтов в результатах поиска, в первую очередь, как раз по коммерческим запросам.

Появление этой группы факторов ранжирования внесло существенные изменения в принципы продвижения сайтов агентствами. Теперь клиенты заинтересованы

в том, чтобы специалисты не просто закупали ссылки, а работали над повышением коммерческого потенциала сайта, делали его удобнее и т. д. В то же время, точное влияние тех или иных коммерческих факторов на ранжирование сайта (страницы) в поисковой выдаче «Яндекс», естественно, не раскрывает. Алгоритмы постоянно совершенствуются, а влияние отдельных факторов может существенно зависеть от тематики и/или частоты запроса, а также от предполагаемых «Яндексом» намерений пользователя, задавшего запрос. Соответственно, качественное продвижение стало требовать более детального понимания принципов, которыми руководствуется поисковая система при оценке коммерческих сайтов, причем важно не только как и какие свойства сайта учитываются при ранжировании, но и как распределяется вес коммерческих, текстовых и ссылочных факторов.

В настоящее время 1-2 раза в месяц собираются результаты поисковой выдачи по более чем 150 запросам, а полученные несколько тысяч сайтов оцениваются по более чем 200 параметрам. Примерно две трети параметров определяются автоматически, а оставшаяся треть (включая многие из «коммерческих» параметров) оценивается специально обученными специалистами — ассессорами. Факторы ранжирования изучаются не только для Яндекса, но и для Google.

Кроме общей аналитической работы (полные результаты которой, в отличие от данных анализаторов качества поиска на AnalyzeThis.ru, мы, к сожалению, не раскрываем) проводится целенаправленный аудит сайтов наших клиентов по тем же методикам, с привлечением наших ассессоров.

Количество исследуемых тематик и параметров постепенно увеличивается, растет также и арсенал статистических методов, которые мы используем. Сейчас мы можем делать выводы о «силе» факторов ранжирования гораздо более уверенно, чем осенью 2014 года,

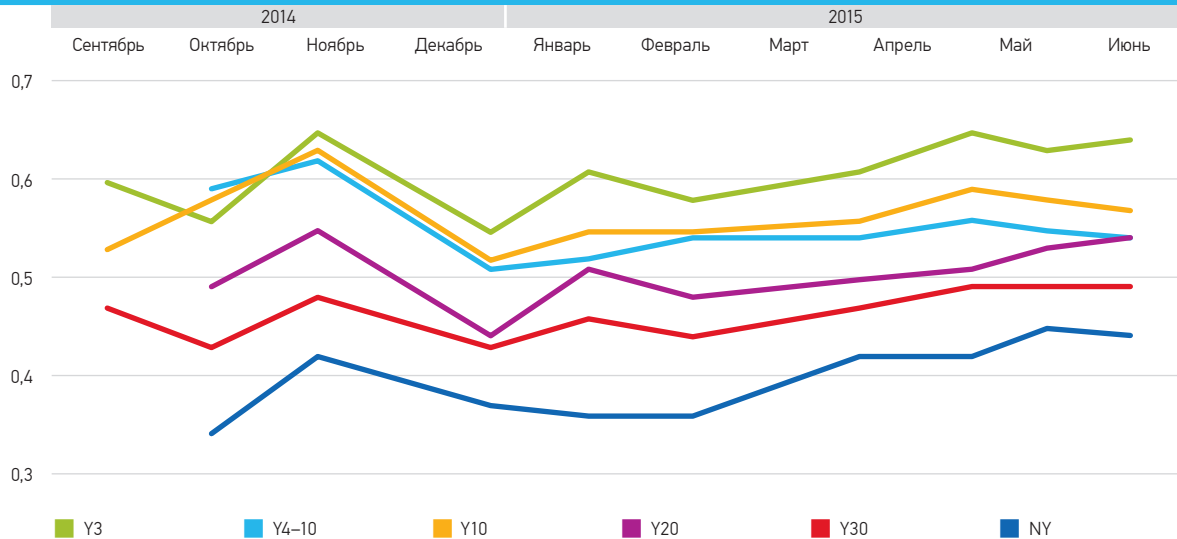
когда впервые рассказали о проекте на конференции «Интернет и бизнес. Россия» (IBC Russia). Мы также наблюдаем факторы ранжирования в динамике, с шагом приблизительно в месяц.

ПО РЕЗУЛЬТАТАМ НАШИХ НАБЛЮДЕНИЙ МОЖНО ДОСТАТОЧНО УВЕРЕННО СДЕЛАТЬ ДВА ВЫВОДА

Во-первых, на ранжирование страницы сайта влияют не только ее собственные характеристики, но и параметры, относящиеся к сайту в целом. В принципе, в этом нет ничего особенно нового — например, известно, что индекс цитирования сайта (домена) в целом влияет на ранжирование его страниц. Однако сейчас мы можем с уверенностью утверждать, что это относится и к многим коммерческим факторам: «Яндекс» активно учитывает в ранжировании страницы коммерческие параметры, значения которых вычисляются им для сайта в целом. Например, наличие отзывов на сайте в целом скоррелировано с позицией, занимаемой страницей сайта в выдаче, сильнее, чем наличие отзывов на этой конкретной странице.

Во-вторых, для каждого параметра полезно оценивать не одну корреляцию, а две. Параметр может быть связан с попаданием сайта (страницы) в «широкий топ» «Яндекса» (сейчас мы оцениваем попадание в топ-30), но не влиять на попадание в десятку и вообще на позицию внутри топ-10. И наоборот, в некоторых случаях сильная корреляция наличия параметра с позицией внутри топа наблюдается при полном отсутствии корреляции с попаданием в топ-30. Разумеется, есть параметры, которые скоррелированы и с попаданием в топ, и с позицией внутри него — как есть и те, для которых никаких значимых корреляций вообще не обнаружено. То же самое верно и для Google — хотя и наборы параметров, для которых наблюдаются корреляции, и тип корреляции (с попаданием в топ

Рисунок 42.



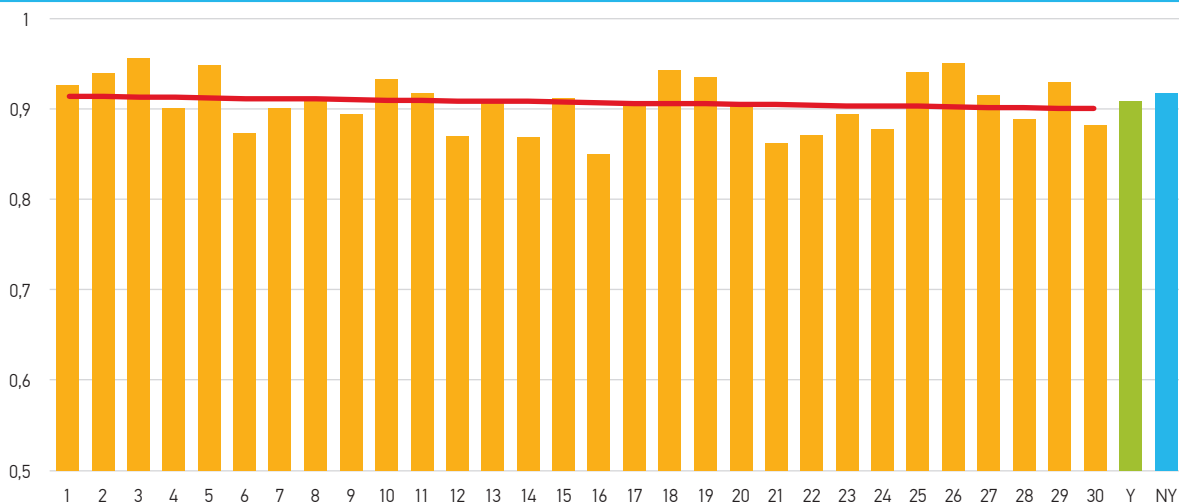
или с позицией в нем), и ее сила для разных поисковиков может быть очень разной (это представляется нам косвенным подтверждением актуальности наших изысканий).

Среди коммерческих свойств сайтов группа факторов, связанных с представлением контактной информации и способами взаимодействия с клиентом, является одной из важнейших. Анализируя их, поисковые машины не только пытаются определить, насколько пользователю будет удобно воспользоваться услугами компании, стоящей за сайтом, но и оценивают «серьезность» и солидность самой компании, ее готовность вкладываться в сервис. Один из исследуемых факторов этой группы — наличие возможности заказать обратный звонок. По нашим данным, это свойство всегда было скоррелировано как с попаданием в топ-30, так и с позицией. На графике

(рис. 42) показаны доли сайтов со свойством в топ-3, среди сайтов на 4-10 позициях, во второй и третьей десятках выдачи Яндекса и в сайтах, найденных по тем же запросам в топах других поисковых систем. Легко видеть, что взаимное расположение точек практически не меняется с ноября, и что чем ближе сайт к топу Яндекса, тем выше вероятность, что на нем присутствует возможность «заказать обратный звонок» (из чего можно сделать осторожный вывод, что желающим попасть в топ Яндекса целесообразно предусмотреть у себя на сайте такую возможность, независимо от ее прямой коммерческой целесообразности).

Другая важная группа факторов — представление товара и удобство работы с ним. В эту группу входят, например, «наличие на сайте характеристик товара» и «подбор по параметрам».

Рисунок 43.



72

Первое свойство присутствует у подавляющего большинства сайтов, причем независимо от поисковика, где они были найдены, поэтому рассчитывать на серьезную корреляцию любого вида не приходится. На картинке (рис. 43) — процент сайтов продавцов и производителей, имеющих данное свойство, по позициям (данные за июнь 2015).

И действительно, ни один из наших статистических методов не показывает никакой корреляции. Зато у второго свойства (возможность «подбора по пара-

метрам») корреляция очень сильная — хотя она и не была такой в самой первой, октябрьской выкачке. Ниже графики распределения по позициям (рис. 44) и динамики (рис. 45): легко видеть, что разрыв между топ-3 Яндекса и сайтами, не попавшими в Яндекс, составляет около 15 процентных пунктов.

Рисунок 44.

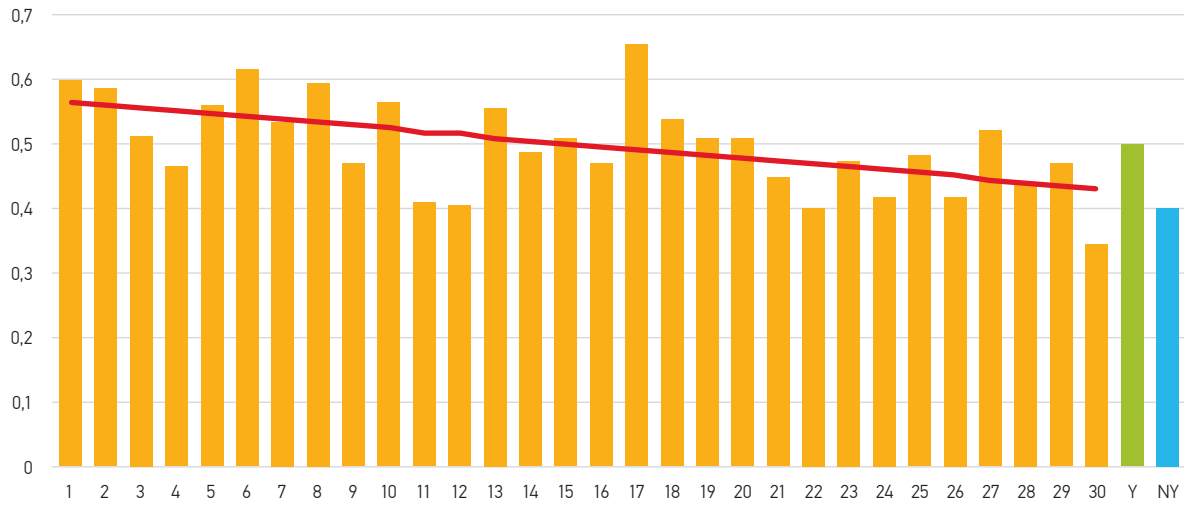


Рисунок 45.

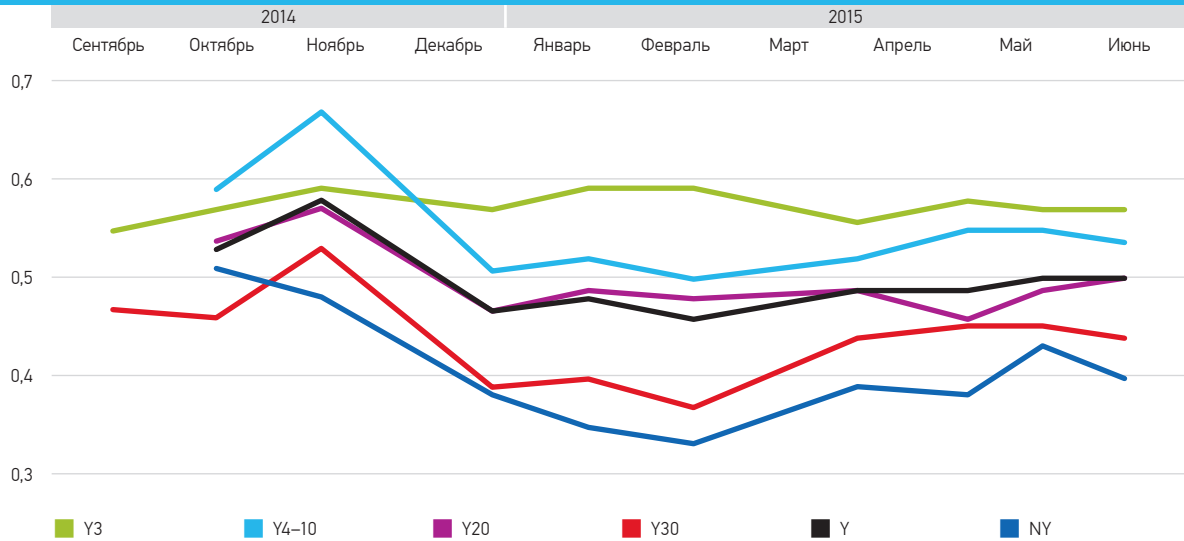
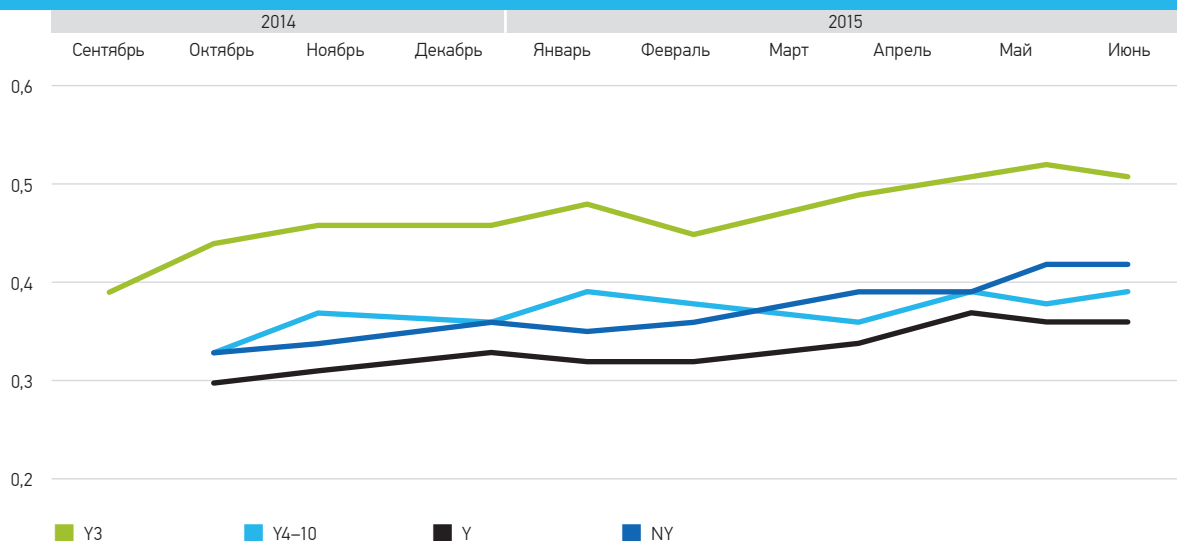


Рисунок 46.



74

Наконец, проиллюстрируем различие в корреляциях «с попаданием в Топ» и «с номером позиции» на примере наличия видео на сайте — свойства, про которое в SEO-среде общеизвестно, что оно «важно». За все время наблюдений доля сайтов с видео в топ-3 стабильно выше доли на позициях 4—10, которая в свою очередь ощутимо выше средней по топ-30. В то же время, доля сайтов с видео вне топ-30 Яндекса иногда даже выше, чем на 4—10 позициях (рис. 46).

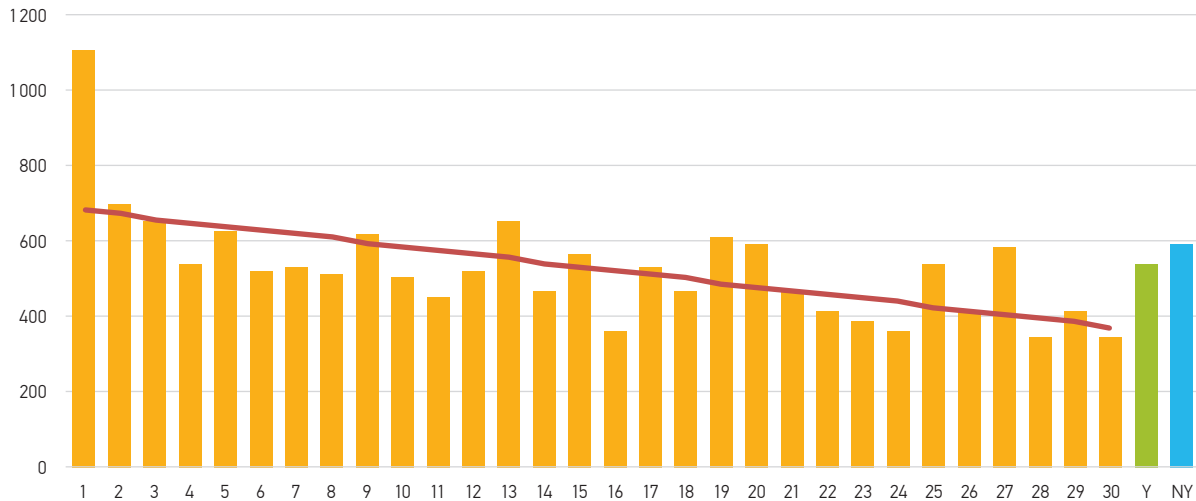
РАЗУМЕЕТСЯ, ИССЛЕДОВАНИЯ ЛАБОРАТОРИИ НЕ ОГРАНИЧИВАЮТСЯ КОММЕРЧЕСКИМИ ФАКТОРАМИ.

Наши исследования показывают, что текстовые факторы, с которых когда-то и начиналось поисковое ранжирование, и сейчас играют очень важную роль —

их недооценка может оказаться критичной, прежде всего, для попадания в топ. Судя по всему, значение текстовых факторов выросло в связи с попытками поисковых машин (и прежде всего Яндекса) бороться с накруткой ссылочных факторов. Изучение текстовых факторов принесло некоторые сюрпризы — например, оказалось, что легендарная функция ранжирования Окари VM25 (связанная с «плотностью» слов запроса в тексте), скорее всего, в настоящее время в ранжировании Яндекса не задействована. Во всяком случае, никаких корреляций между факторами, использующими эту функцию, и позицией в Яндексе обнаружить не удалось.

Особую роль в исследованиях Лаборатории занимают ссылочные факторы. Учитываются как общие ссылочные факторы (количество ссылок/ссылающихся доменов на домен и конкретную страницу, ТИЦ, PageRank),

Рисунок 47.



так и «текстовые ссылочные»: количество ссылок, содержащих слова запроса (точный запрос, достаточное количество слов запроса и т. п.). Несмотря на то, что мы имеем дело только с коммерческими запросами, для которых Яндекс, как известно, уже более года назад «отключил» ссылочное ранжирование, большинство ссылочных факторов остаются одними из самых сильных. Более того, осенью 2014 года мы специально проверяли, влияют ли «покупные» ссылки — и их «работоспособность» блестяще подтвердилась.

Сказанное выше вовсе не означает, что в ссылочном ранжировании Яндекса в 2014 году ничего не изменилось и слова про отключение ссылочного ранжирования для коммерческих запросов не имеют под собой реального основания. Скорее всего, какие-то из многочисленных ссылочных факторов были отклю-

чены, а оставшиеся влияют на выдачу «косвенно», наряду с коммерческими, текстовыми и поведенческими факторами. В целом значение ссылочных факторов, хотя и остается достаточно высоким, перестало быть преобладающим — грубо говоря, теперь уже далеко не всегда можно «продвинуть сайт в топ» исключительно за счет покупки ссылок.

По понятным причинам с мая 2015 года Лаборатория изучения факторов ранжирования уделяет особое внимание исследованию «Минусинска» — нового алгоритма Яндекса, предусматривающую наказание сайтам за покупку ссылок. На основании предварительных данных уже можно констатировать, что новый алгоритм, вопреки риторике Яндекса, все-таки не исключает ссылочные факторы из ранжирования.

Разумеется, значение ссылочных факторов не могло не уменьшиться за счет «перемешивания» выдачи, когда пессимизированные сайты с большим ссылочным весом провалились на более низкие позиции или вообще выпали из топа. Тем не менее, ссылочные факторы продолжают оставаться достаточно сильными. На графике (рис. 47) — средние значения ТИЦ по позициям в топ-30 Яндекса на июнь 2015.

Если не учитывать сайты, попавшие под «Минусинск», корреляция среднего ТИЦ с позицией в Яндексе становится еще более выраженной, а наклон аппроксимирующей прямой — практически таким же, каким он был до «Минусинска».

МОСКВА, 2015
«АШМАНОВ И ПАРТНЕРЫ»
ВСЕ ПРАВА ЗАЩИЩЕНЫ
WWW.ASHMANOV.COM